PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/56, 15/62, 15/80
// C12N 15/55

A1

(11) International Publication Number: WO 93/12237

(43) International Publication Date: 24 June 1993 (24.06.93)

(21) International Application Number: PCT/EP92/02896

(22) International Filing Date: 9 December 1992 (09.12.92)

(30) Priority data:
9102051 9 December 1991 (09.12.91) NL

(71) Applicant *(for AU CA GB IE MW only)*: UNILEVER PLC [GB/GB]; Unilever House, Blackfriars, London EC4P 4BQ (GB).

(71) Applicant *(for all designated States except AU CA GB IE MW US)*: UNILEVER N.V. [NL/NL]; P.O. Box 137, NL-3000 DK Rotterdam (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : GOUKA, Robertus, Johannes [NL/NL]; Drenthe Plantsoen 48, NL-2545 BL Den Haag (NL). VAN DEN HONDEL, Cornelis, Antonius, M., J. [NL/NL]; Waterlelie 124, NL-2804 PZ Gouda (NL). MUSTERS, Wouter [NL/NL]; Wipperspark 138, NL-3141 RD Maassluis (NL). STAM, Hein [NL/NL]; Griend 72, NL-1112 LG Diemen (NL). VERBAKEL, Johannes, Maria, A. [NL/NL]; Ingeland 9, NL-3155 GC Maasland (NL).

(74) Agent: VAN DER TOORREN, Johannes; Unilever N.V., Patent Division, P.O. Box 137, NL-3130 AC Vlaardingen (NL).

(81) Designated States: AU, BG, CA, CS, FI, HU, JP, KR, MW, NO, PL, RO, RU, SD, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PROCESS FOR PRODUCING/SECRETING A PROTEIN BY A TRANSFORMED MOULD USING EXPRESSION/SECRETION REGULATING REGIONS DERIVED FROM AN *ASPERGILLUS* ENDOXYLANASE II GENE

(57) Abstract

Methods are described for the isolation and characterization of DNA sequences from *Aspergillus niger* var. *awamori* which are involved in the expression and secretion of endoxylanase II (exlA) by said *Aspergillus* mould. A process using these expression and/or secretion regulating regions to direct the production and optionally the secretion of proteins other than endoxylanase II by transformed moulds is provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SK	Slovak Republic
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
UA	Ukraine
US	United States of America
VN	Viet Nam

Title: **Process for producing/secreting a protein by a transformed mould using expression/secretion regulating regions derived from an *Aspergillus* endoxylanase II gene**

Background of the invention

The invention relates to a process for the production and optionally secretion of a protein by means of a transformed mould, into which an expression vector has been introduced with the aid of recombinant DNA techniques known per se, said vector comprising one or more mould-derived expression and/or secretion regulating regions controlling the expression of a gene encoding said protein and optionally controlling the secretion of the protein so produced. Such a process is known from various publications, in which the production of proteins with the aid of transformed moulds is described. Thus, in the non-prior-published patent application PCT/EP 91/01135 (UNILEVER, in the priority year published on 26 December 1991 as WO 91/19782) there is described, **inter alia**, the production of a homologous endoxylanase II protein by a transformed *Aspergillus* strain.

Other ways of producing proteins by transformed moulds, in particular while using promoters originating from *Aspergillus* moulds, are known.

- Ward, M. *et al.*, (GENENCOR, 1990) have described the production by a transformed *Aspergillus niger* var. *awamori* of the milk-clotting enzyme chymosin or its precursor prochymosin. It was concluded that production of a fusion protein in which the prochymosin was connected with its N-terminus to the C-terminus of the *Aspergillus* protein glucoamylase gave a much higher secretion than with production of the prochymosin alone, whereby in both cases the protein was preceded by the glucoamylase signal sequence and under control of the glucoamylase promoter.
- In CA-A-2024448 (ALLELIX BIOPHARMACE) "Recombinant DNA expression construct - containing promoter for use in *Aspergillus*", published on 1 March 1991, the constitutive promoter of the *Aspergillus nidulans* aldehyde dehydrogenase gene and its use for the production of heterologous proteins in a transformed mould is described.

- In EP-A-0436858 (GREEN CROSS CORP.) "Promoter of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene - derived from *Aspergillus orizae*, used in new expression system in yellow-green or black koji mould", published on 17 July 1991, the use of the promoter and terminator of the GAPDH gene in a vector for transforming a mould to produce foreign proteins is described.

and

- In EP-A-0439997 (CIBA GEIGI AG) "*A. niger* pyruvate kinase promoter - used to construct vectors for expression of structural genes in suitable hosts", published on 7 August 1991, the overproduction of a homologous gene product or a heterologous gene product in *A. niger* is described.

Moulds are organisms frequently used in the production of proteins and metabolites. A biotechnologically very important aspect of moulds is that they are capable of very efficient protein production and, if desired, secretion into the medium. It is also possible to grow moulds in a properly controlled way in large bioreactors. The combination of the possibilities of generating fungal biomass cost-effectively by means of fermentation and the high specific expression per cell make moulds exceptionally interesting hosts for the production of both heterologous and homologous proteins. For efficient production of these heterologous and homologous proteins, the use of an efficient promoter effective in moulds is essential. For secretion of a protein into the medium, specific sequences are required that cater for this. In connection with possible toxicity for the mould cell of the protein to be produced, it is also important that the activity of a promoter can be regulated, i.e. turned on at suitable moments, thus an inducible promoter is preferred.

Summary of the invention

The invention is based on the use of a non-prior-published promoter, which is described in more detail below, as well as on the use of other expression and/or secretion regulating regions, such as a terminator, a DNA sequence encoding a signal sequence, and a DNA sequence encoding at least an essential part of a mature endogenous mould protein.

In studies of the expression of proteins in moulds it was found that the enzyme endoxylanase type II (exlA) was efficiently produced after induction of expression of the exlA gene, and was also secreted efficiently into the medium. For production of that protein, the encoding gene was cloned together with its own promoter. In comparison with other mould promoters, the endoxylanase II promoter proved particularly efficient. Expression of the gene encoding the endoxylanase II enzyme (regulated by its own promoter) was found to be efficiently induced with various media components, including wheat bran, xylan and xylose. This induction was found to proceed efficiently in different mould strains (see WO 91/19782, UNILEVER). This provided an opportunity to obtain an efficient inducible promoter as well as other mould-derived expression and/or secretion regulating regions, including transcription terminator signals and secretion signals, which might perhaps be used for the production of heterologous and homologous proteins in moulds. The promoter fragment, terminator fragment and secretion signals of the *Aspergillus niger* var. *awamori* endoxylanase II gene were cloned and subsequently further defined.

The *E. coli* ß-glucuronidase gene (uidA) was used as an example of the production of a heterologous protein in a transformed mould. The promoter and terminator sequences of the *Aspergillus niger* var. *awamori* endoxylanase II gene were used for the construction of an expression vector. With the aid of this expression vector a heterologous gene encoding the *E. coli* protein ß-glucuronidase was expressed in moulds under control of the endoxylanase II (exlA) promoter. By using exlA secretion signals, the heterologous and homologous proteins can also be secreted.

As another example of the use of the exlA promoter and terminator for the production of heterologous proteins, a gene encoding a *Thermomyces lanuginosa* lipase was introduced in the expression vector under the control of exlA regulatory sequences, and used for the production and secretion of *Thermomyces lanuginosa* lipase. NOVO-NORDISK, an enzyme manufacturing company in Denmark, is marketing under the trade name "Lipolase" a lipase derived from *Thermomyces lanuginosa*, but produced by another microorganism. To illustrate that the exlA signal sequence can be used to direct the secretion of proteins other than exlA, a DNA sequence encoding a *Thermomyces lanuginosa* mature lipase amino acid sequence was fused to the exlA signal sequence and placed under the control of the exlA regulatory

sequences in the above mentioned expression plasmid, and secretion of *Thermomyces lanuginosa* lipase was demonstrated.

Of course, the heterologous genes, of which the expression is exemplified in this specification, can be replaced by any DNA sequence encoding a desired protein (coding for enzymes, proteins, etc.) originating from a wide range of organisms (bacteria, yeasts, moulds, plants, animals and human beings) so that the desired protein can be produced by moulds.

Thus in one embodiment of the invention a process is provided for the production in transformed moulds of proteins other than endoxylanase type II using expression regulating sequences derived from the *Aspergillus niger* var. *awamori* endoxylanase II (exlA) gene, such as the promoter or the terminator, or functional derivatives of these regulatory sequences. In another embodiment of this invention, a process is provided by which proteins produced in moulds, if desired, can be secreted in the medium by making use of the DNA sequence encoding the signal sequence, in particular the pre-sequence or prepro-sequence, of the *Aspergillus niger* var. *awamori* endoxylanase II gene or functional derivatives of these sequences. Finally the invention also provides a process for producing a protein in which a vector is used comprising at least an essential part of the DNA sequence encoding the mature endoxylanase II protein, because it is known that in moulds an improved secretion of a heterologous protein can be obtained by initially producing it as a fusion protein comprising part of an endogenous mould protein (see also the Ward, M. *et al.* / GENENCOR reference mentioned above).

Brief description of the Figures and Tables

Fig. 1 shows the DNA sequence of the ca 2.1 kb *Pst*I-*Pst*I fragment of *Aspergillus niger* var. *awamori* present in the plasmid pAW14B, which fragment contains a gene coding for an endoxylanase II, indicated as the exlA gene. The translation start and the stop codon are doubly underlined. The 49 bp intron is underlined. The N-terminal end of the mature protein is indicated. The amino acid sequence of the protein (both of the pre(pro) form and of the mature protein) is indicated using the one-letter code.

Fig. 2 shows the restriction map of the genomic DNA region of *Aspergillus niger* var. *awamori*, comprising the exlA gene cloned in the phages lambda 1 and lambda 14. The used abbreviations stand for: S: *Sal*I; E: *Eco*RI; H: *Hin*dIII; P: *Pst*I; P*: *Pst*I; B: *Bam*HI; S#: *Sal*I site originating from the polylinker of lambda-EMBL3; and D: *Sau*3A. The solid bar indicates a 1.2 kb *Pst*I*-*Bam*HI fragment hybridizing with Xyl06. P* and *Pst*I* symbols are used to distinguish the two *Pst*I sites present.

Fig. 3 shows the plasmid PAW14B obtained by insertion of the 5.3 kb *Sal*I fragment comprising the exlA gene of *Aspergillus niger* var. *awamori* in the *Sal*I site of pUC19.

Fig. 4 shows the plasmid pAW15-1 obtained by displacing the *Bsp*HI-*Afl*II fragment comprising the exlA open reading frame in pAW14B with a *Nco*I-*Afl*II fragment comprising the *E. coli* uidA coding sequence. Thus, plasmid pAW15-1 comprises the *E. coli* uidA gene under the control of the *A. niger* var. *awamori* promoter and terminator.

Fig. 5 shows plasmid pAW15-7 obtained by inserting a 2.6 kb *Not*I fragment comprising the *E. coli* hygromycin resistance gene controlled by the *A. nidulans* gpdA promoter and the *A. nidulans* trpC terminator in the *Eco*RI site of pAW15-1.

Fig. 6 shows plasmid pAWTL1 obtained by displacing the *Bsp*HI-*Afl*II fragment comprising the exlA open reading frame in pAW14B with a *Bsp*HI-*Afl*II fragment comprising a nucleotide sequence encoding the *T. lanuginosa* lipase together with its own pre-pro-sequence. Thus, plasmid pAWTL1 comprises the *T. lanuginosa* lipase gene together with its own pre-pro-sequence encoding region under the control of the *A. niger* var. *awamori* promoter and terminator.

Fig. 7 shows plasmid pAWTL2 obtained by displacing the *Nru*I-*Afl*II fragment comprising the region encoding the mature exlA protein in pAW14B with a *Nru*I-*Afl*II fragment comprising a nucleotide sequence encoding the mature part of the *T. lanuginosa* lipase. Thus, plasmid pAWTL2 comprises the *T. lanuginosa* lipase gene

fused to the exlA pre-pro-sequence encoding region under the control of the *A. niger* var. *awamori* promoter and terminator.

Fig. 8 shows plasmid pTL1 comprising a nucleotide sequence encoding the *T. lanuginosa* lipase together with its own pre-pro-sequence under the control of the *A. niger* gpdA promoter and the *A. nidulans* trpC terminator inserted in the polylinker of pUC18. The region encoding the pre-pro-sequence of the *T. lanuginosa* lipase is indicated by "ss"

Fig. 9 shows the sequence comprising the open reading frame encoding the *T. lanuginosa* lipase as it is contained within plasmid pTL1. The N-terminal end of the mature protein is indicated.

Table A shows various probes derived from the N-terminal amino acid sequence of the endoxylanase II protein. These probes were used for the isolation of the exlA gene, see item 1.1 of Example 1.
The number of oligonucleotides present in the "mixed" probe is indicated in brackets; this number is obtained by including 1, 2, 3 or 4 different bases in every third position, depending on the number of codons for an amino acid. In Xyl04 nucleotides were selected on the basis of the hybridization G-C and G-T and/or on the basis of the preferred codons in *Aspergillus niger* glucoamylase. In Xyl05 and Xyl06 not all of the possibly occurring bases are introduced at the third position of the codons in order not to obtain more than 256 oligonucleotides in the mixture. The sequence of the oligonucleotides is complementary to that of the coding strand of the DNA, which resembles the corresponding mRNA.

Xyl01: a mixture of 256 oligos having a length of 23 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 5-12.

Xyl04: an oligo having a length of 47 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 2-17.

Xyl05: a mixture of 144 oligos having a length of 23 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 10-17.

Xyl06: a mixture of 256 oligos having a length of 47 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 2-17.

Table B shows various single-stranded subclones of lambda 1 and lambda 14 fragments, which were used for determination of the sequence of the exlA gene, see item 1.2 of Example 1.

Table C shows the results of *E. coli* ß-glucuronidase production by non-transformed and transformed strains of the mould *Aspergillus niger* var. *awamori*, see item 2.2 of Example 2.

Table D shows that functional lipase was produced and secreted after induction of the exlA promoter by xylose, and that the secretion of a heterologous (*Thermomyces lanuginosa*) mature protein was directed in *Aspergillus niger* var. *awamori* by using either the exlA signal sequence (see AWLPL2-2) or the *Thermomyces lanuginosa* signal sequence (see AWLPL1-2), see item 3.2 of Example 3.

Table E shows various nucleotide sequences of oligonucleotides used in constructions described in Examples 1-3, see items 1.4, 2.1, 3.1.1, 3.1.2, 3.1.3, and 3.1.4. The sequence listing numbers refer to the listings provided in the official format.

Detailed description of the invention

Since the endoxylanase II gene is expressed and the resulting protein is secreted very efficiently under appropriate cultivation conditions by *Aspergillus niger* var. *awamori*, the present invention is directed in particular to the cloning of the regulatory regions of the *Aspergillus niger* var. *awamori* endoxylanase II (exlA) gene, such as the promoter sequence, terminator sequence and signal sequence, and using these components for the development of a process for the production of proteins in moulds. The invention therefore relates generally to a process making use of a nucleic acid sequence derivable from a mould and comprising at least a regulatory region derivable from a gene encoding a polypeptide having endoxylanase II activity. Said nucleic acid sequence can be combined with nucleic acid sequences encoding other

homologous or heterologous genes to bring these genes under the control of at least one exlA regulatory sequence.

"Nucleic acid sequence" as used herein refers to a polymeric form of nucleotides of any length, both to ribonucleic acid (RNA) sequences and deoxyribonucleic acid (DNA) sequences. In principle this term refers to the primary structure of the molecule. Thus this term includes both single and double stranded DNA, as well as single stranded RNA and modifications thereof.

In general the term "protein" refers to a molecular chain of amino acids with a biological activity and does not refer to a specific length of the product and if required can be modified in vivo or in vitro. This modification can for example take the form of amidation, carboxylation, glycosylation, or phosphorylation; thus inter alia peptides, oligopeptides and polypeptides are included. In this specification both terms, polypeptide and protein, are used as synonyms unless a specific meaning is clear from the context.

The invention also relates to the use of a vector containing the nucleic acid sequences as described for the production of proteins other than *Aspergillus niger* var. *awamori* endoxylanase II (exlA) and also relates to the use of micro-organisms containing said vectors or nucleic acid sequences for producing said proteins.

The invention is also directed at the use of modified sequences of the aforementioned nucleic acid sequences according to the invention for the production of proteins other than *Aspergillus niger* var. *awamori* endoxylanase II (exlA), said modified sequences also having regulatory activity. The term "a modified sequence" covers nucleic acid sequences having the regulatory activity equivalent to or better than the nucleic acid sequence derivable from a mould and comprising at least a regulatory region derivable from a gene encoding a protein having endoxylanase II activity. Such an equivalent nucleic acid sequence can have undergone substitution, deletion or insertion or a combination of the aforementioned of one or more nucleotides resulting in a modified nucleic acid sequence without concomitant loss of regulatory activity occurring. Processes for the production of proteins other than *Aspergillus niger* var. *awamori* endoxylanase II (exlA) using such modified nucleic acid sequences fall within the scope of the present invention. In particular processes for the production of proteins other than *Aspergillus niger* var. *awamori* endoxylanase II (exlA) using

modified sequences capable of hybridizing with the non-modified nucleic acid sequence and still maintaining at least the regulatory activity of the non-modified nucleic sequence fall within the scope of the invention.

The expression "functional derivatives" used in the claims refers to such modified sequences.

The term "a part of" covers a nucleic acid sequence being a subsequence of the nucleic acid sequence derivable from a mould and comprising at least a regulatory region derivable from a gene encoding a polypeptide having endoxylanase II activity.

In particular the invention is directed at a process using a nucleic acid sequence derivable from a mould of the genus *Aspergillus*. A suitable example of a mould from which a nucleic acid sequence according to the invention can be derived is an *Aspergillus* of the species *Aspergillus niger*, in particular *Aspergillus niger* var. *awamori*. In particular the strain *Aspergillus niger* var. *awamori* CBS 115.52 (ATCC 11358) is eminently suitable for deriving a nucleic acid sequence according to the invention.

Preferably the nucleic acid sequence for use in a process according to the invention comprises at least a promoter as regulatory region. The nucleic acid sequence for use in a process according to the invention can also comprise an inducer or enhancer sequence enabling a higher level of expression of any nucleic acid sequence operably linked to the promoter. It is also possible for the nucleic acid sequence for use in a process according to the invention to comprise a termination signal as regulatory region. The nucleic acid sequence for use in a process according to the invention can comprise one or more regulatory regions. A nucleic acid sequence for use in a process according to the invention can comprise solely the promoter as regulatory region or a combination thereof with an enhancer or other functional elements. A nucleic acid sequence for use in a process according to the invention can also further comprise terminator sequences, although these are not always required for efficient expression of the desired expression product.

According to a further embodiment of the invention a nucleic acid sequence for use in a process according to the invention can further comprise a sequence encoding a secretory signal necessary for secreting a gene product from a mould. This will be preferred when intracellular production of a desired expression product is not sufficient and extracellular production of the desired expression product is required.

Secretory signals comprise the prepro- or pre-sequence of the endoxylanase II gene for example. A secretory signal derivable from the endoxylanase II gene of an *Aspergillus* mould is particularly favoured. The specific embodiment of the nucleic acid sequence used in a process according to the invention will however depend on the goal that is to be achieved upon using a process according to the invention. "Signal sequence" as used herein generally refers to a sequence of amino acids which is responsible for initiating export of a protein or polypeptide chain. A signal sequence, once having initiated export of a growing protein or polypeptide chain, can be cleaved from the mature protein at a specific site. The term also includes leader sequences or leader peptides. The preferred signal sequence herein is the deduced signal sequence from the *Aspergillus niger* var. *awamori* endoxylanase II gene given in Fig. 1.

With the help of DNA oligonucleotides deduced from protein sequence analysis of endoxylanase II from *Aspergillus niger* var. *awamori* chromosomal DNA fragments comprising the entire endoxylanase II (exlA) gene of *Aspergillus niger* var. *awamori* including the regulatory regions, such as the promoter, the signal sequence and the termination sequence have been isolated from a genomic library. The regulatory regions of the endoxylanase II (exlA) gene have been used for the production and, if desired, secretion of proteins other than endoxylanase II, e.g. heterologous proteins, by *Aspergillus niger* var. *awamori*. The invention is therefore in particular directed at a process in which one or more of the regulatory regions of the *Aspergillus niger* var. *awamori* endoxylanase II gene or equivalent nucleic acid sequences are used for the production of proteins other than endoxylanase II in *Aspergillus niger* var. *awamori*. The term "equivalent nucleic acid sequence" has the same meaning as given above for "a modified nucleic acid sequence".

In the Examples given below the expression and secretion potential of the obtained exlA promoter and the exlA signal sequences have been tested by constructing new vectors for expression of a heterologous ß-glucuronidase gene and the production and secretion of a heterologous lipase in *Aspergillus*. The resulting constructs were tested in *Aspergillus niger* var. *awamori*.

Thus in a general form the invention provides a process for the production and optionally secretion of a protein different from the endoxylanase type II protein ex *Aspergillus niger* var. *awamori* by means of a transformed mould, into which an expression vector has been introduced with the aid of recombinant DNA techniques known per se, said vector comprising mould-derived expression and/or secretion regulating regions, in which process at least one of said expression and/or secretion regulating regions is selected from (1) the expression and secretion regulating regions of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and (2) functional derivatives thereof also having expression and/or secretion regulating activity.

In a preferred embodiment of the invention the selected expression regulating region is a promoter and said vector comprises a gene encoding said protein under control of said promoter, the latter being selected from (1) the promoter of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and (2) functional derivatives thereof also having promoter activity. More preferably said promoter is equal to the promoter present on the 5' part upstream of the exlA gene having a size of about 2.5 kb located between the *Sal*I restriction site at position 0 and the start codon ATG of the exlA gene in plasmid pAW14B, in particular said promoter comprises at least the polynucleotide sequence 1-350 according to Figure 1.
This promoter can be induced by wheat bran, xylan, or xylose, or a mixture of any combination thereof, present in a medium in which the transformed mould is incubated, whereby the use of xylose as inducing agent is preferred.

In another preferred embodiment of the invention the selected expression regulating region is a terminator and said vector comprises a gene encoding said protein followed by said terminator, the latter being selected from (1) the terminator of the

endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and (2) functional derivatives thereof also having terminator activity. Preferably, said terminator is equal to the terminator present on the 3' part downstream of the exlA gene having a size of about 1.0 kb located right downstream of the stop codon (TAA) of the exlA gene in plasmid pAW14B.

A further embodiment of the invention is a process for the production and secretion of a protein different from the endoxylanase type II protein ex *Aspergillus niger* var. *awamori* by means of a transformed mould, in which process the selected secretion regulating region is a DNA sequence encoding a signal sequence and said vector comprises a gene encoding said protein preceded by said DNA sequence encoding a signal sequence, the latter being selected from (1) the DNA sequence encoding the signal sequence of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and (2) functional derivatives thereof also directing secretion of the protein. Preferably, the gene (1) encoding said protein is also preceded by at least an essential part of a DNA sequence (2) encoding the mature endoxylanase II protein, whereby said DNA sequence (2) is present between said DNA sequence encoding a signal sequence (3) and the gene (1). A preferred signal sequence is the signal sequence encoded by polynucleotide 351-431 of the DNA sequence given in Figure 1, which polynucleotide precedes the DNA sequence in plasmid pAW14B encoding the mature exlA polypeptide.
Summarizing, in a process for producing a protein according to the invention the vector used for transforming a mould can comprise an exlA-derived promoter as hereinbefore described or a exlA-derived terminator as hereinbefore described or a exlA-derived signal sequence as hereinbefore described or at least an essential part of the exlA structural gene, or any combination of these expression and/or secretion regulating regions.

The invention is illustrated with the following Examples without being limited thereto. All techniques used for the manipulation and analysis of nucleic acid materials were performed essentially as described in Sambrook et al. (1989), except where indicated otherwise.

Example 1 **Cloning and characterization of the endoxylanase II gene (exlA) and associated regulating sequences of *Aspergillus niger* var. *awamori***

1.1 Isolation of the *Aspergillus niger* var. *awamori* exlA gene

In order to isolate the exlA gene from chromosomal DNA of *Aspergillus niger* var. *awamori* different probes were synthesized consisting of mixtures of oligonucleotides (Table A). The composition of these mixtures was derived from the N-terminal amino acid sequence of purified endoxylanase II protein.

By means of Southern blot analysis it was established that in digests of chromosomal DNA - under stringent conditions - only one band hybridizes with the probes used. In the *Eco*RI, *Sal*I and *Bam*HI digest of *Aspergillus niger* var. *awamori* DNA one band of respectively 4.4, 5.3 and 9.5 kb hybridizes with both Xyl01, Xyl04 and Xyl06. With Xyl05 no clear signal was found at 41°C. On the basis of this result a gene bank of *Aspergillus niger* var. *awamori* DNA was hybridized at 65°C with the oligonucleotide mixture Xyl06 as a probe. Of the 65000 tested plaques (corresponding to 32 times the genome) three plaques (lambda 1, 14 and 63) hybridized with this probe. After hybridization of digests of lambda 1 and lambda 14 DNA with Xyl06 a hybridizing band of >10 kb was found in the *Eco*RI digest of lambda 1. The size of the hybridizing band in the lambda 14 and the chromosomal *Eco*RI digest was 4.4. kb. In the *Sal*I digest of lambda 1 a 4.6 kb band hybridizes; in the *Sal*I digest of lambda 14 this is, like in chromosomal DNA, a 5.3 kb band. Also a 1.2 kb *Pst*I-*Bam*HI fragment (Fig. 2) hybridizes with Xyl06. On the basis of restriction patterns with different enzymes and cross-hybridization of lambda 1 and lambda 14 digests with the 5.3 kb

Amino acid sequence																
1			5					10						15		
Ser	Ala	Gly	Ile	Asn	Tyr	Val	Gln	Asn	Tyr	Asn	Gly	Asn	Leu	Gly	Asp	Phe

Probe	Bases	Amino acids	(number of oligonucleotides)													
Base sequence 3' - 5'																
Xyl01	23	5-12					(256)									
			TTA	ATA	CAX	GTT	TTA	ATA	TTA	CC						
				G	G		C	G	G	G						
Xyl04	47	2-17					(1)									
	CGG	CCG	TAG	TTG	ATG	CAG	GTC	TTG	ATG	TTG	CCG	TTG	GAC	CCG	CTG	AA
Xyl05	23	10-17						(144)								
								ATG	TTG	CCA	TTA	AAX	CCA	CTG	AA	
										G	G	G	G			
										C			C			
Xyl06	47	2-17					(256)									
	CGG	CCG	TAG	TTG	ATG	CAG	GTC	TTG	ATG	TTG	CCG	TTG	GAG	CCG	CTG	AA
	C	C	C			C	T				C		C	C		

X= A, G, C or T

Table A **Probes derived from the N-terminal amino acid sequence of the endoxylanase II protein**

The number of oligonucleotides present in the "mixed" probe is indicated in brackets; this number is obtained by including 1, 2, 3 or 4 different bases in every third position, depending on the number of codons for an amino acid. In Xyl04 a G was selected on the basis of the hybridization G-C and G-T and/or on the basis of the preferred codons in *Aspergillus niger* glucoamylase. In Xyl05 and Xyl06 not all possibly occurring bases have been introduced at the third position of the codons in order to obtain no more than 256 oligonucleotides in the mixture. The sequence of the oligonucleotides is complementary to that of the coding strand.

Xyl01: a mixture of 256 oligos having a length of 23 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 5-12.

Xyl04: an oligo having a length of 47 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 2-17.

Xyl05: a mixture of 144 oligos having a length of 23 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 10-17.

Xyl06: a mixture of 256 oligos having a length of 47 deoxynucleotides the sequence of which is complementary to the part of the coding strand coding for the amino acids 2-17.

*Sal*I fragment of lambda 14 it was confirmed that these lambda's contained overlapping fragments of the genome of *Aspergillus niger* var. *awamori*. Also homologous hybridization of total induced RNA with respectively lambda 1, lambda 14 and the 5.3 kb *Sal*I fragment of lambda 14 confirmed the presence of exlA sequences on these lambda's. Hybridization was found with a xylan-induced MRNA of ca. 1 kb. The size thereof corresponds to that of the mRNA molecule hybridizing with Xyl06.

1.2 Subcloning of the *Aspergillus niger* var. *awamori* exlA gene

The *Sal*I fragments hybridizing with Xyl06 of respectively lambda 1 (4.6 kb) and lambda 14 (5.3 kb) were cloned in two orientations in the *Sal*I site of pUC19, which resulted in respectively plasmid pAW1 (A and B) and plasmid pAW14 (A and B, see Fig. 3). The 1.2 kb *Pst*I*-*Bam*HI fragment hybridizing with Xyl06 and the adjacent 1.0 kb *Bam*HI-*Pst*I fragment from respectively pAW14A and pAW1A were subcloned into M13mp18 and M13mp19 cut with *Bam*HI and *Pst*I, resulting in the m18/m19 AW vectors of Table B.

Fragment	Resulting vectors
pAW 1A BamHI-*Pst*I* (1.2 kb)	m18AW 1A-1 / m19AW 1A-1
pAW14A BamHI-*Pst*I* (1.2 kb)	m18AW14A-1 / m19AW14A-1
pAW 1A *Pst*I-BamHI (1.0 kb)	m18AW 1A-2 / m19AW 1A-2
pAW14A *Pst*I-BamHI (1.0 kb)	m18AW14A-2 / m19AW14A-2

Table B Single-stranded subclones of lambda 1 and lambda 14 fragments

1.3 Determination of the transcription direction of the exlA gene.

The transcription direction of the exlA gene was established by means of spot blot hybridization of ss-DNA of respectively m18AW14A-1 and m19AW14A-1 with Xyl06. It was found that ss-DNA of m19AW14A-1 (5'-*Pst*I*-*Bam*HI-3') hybridizes with this probe. Because the sequence of Xyl06 is equal to that of the non-coding strand, m19AW14A-1 contains the coding strand. On the basis thereof the transcription direction shown in Fig. 2 was determined. This direction is confirmed by the results of a primer extension experiment.

1.4 Identification of the exlA gene

The DNA sequence of a part of the promoter region was determined by sequence analysis of pAW14B with Xyl06 as a primer (5' part of the gene). In this region a primer Xyl11 (see Table E) was selected, with which the DNA sequence of complementary strand of m18AW14A-1 and m18AW1A-1 was determined. The results showed that these vectors contained a DNA sequence which was substantially equal to that of Xyl06, while the amino acid sequence derived from the base pair sequence was identical with the N-terminal amino acid sequence of the mature endoxylanase II protein. Thus the cloning of at least the 5' end of the exlA gene had been proven. The presence of the entire exlA gene in the vectors pAW14 and pAW1 seemed plausible on the basis of the position of the 5' end of the gene on the *Sal*I fragments (Fig. 2) and the size of the exlA mRNA (ca. 1 kb).

1.5 Sequence analysis

The nucleotide sequence of the exlA gene and surrounding regions was established in two directions in both the m13AW14 and the m13AW1 subclones by means of the dideoxy procedure (Sanger et al., 1977). The sequence around the *Bam*HI site located downstream of the *Pst*I* site (Fig. 2) was established by sequence analysis of double-stranded pAW14 and pAW1 DNA. Compressions were cleared up by using dITP instead of dGTP. In the independent clones lambda 1 and lambda 14 an identical exlA sequence was established. The complete nucleotide sequence of the 2.1 kb *Pst*I*-*Pst*I fragment comprising the entire pre(pro) endoxylanase II gene and the promoter and terminator sequences of the endoxylanase II gene is shown in Fig. 1. The mature endoxylanase II protein is preceded by a leader peptide of 27 amino acids. A

predicted recognition site for the signal peptidase is present between the alanine residues at the positions 16 and 17 (..-T-A-F-A-I-A-P-V-..) (Van Heijne, 1986). From the length of the leader peptide it can be derived that in the protein a second processing site is present. Cleavage of the bond between Arg (27) and Ser (28) presumably is performed by a KEX2-like endoprotease (Fuller et al., 1988).

1.6 Localization of the intron

In the exlA gene the presence of an intron of either 49 or 76 bp (581-629 or 581-656, see Fig. 1) was predicted on the basis of the presence of sequences corresponding to "donor" and "acceptor" sites of introns in aspergilli. Definite proof of the absence of a 76 bp intron was obtained by isolation of an endoxylanase II derived peptide with the sequence Tyr-Ser-Ala-Ser-Gly... This peptide can only be localized in the protein starting from nucleotide position 652 (see Fig. 1). Therefore, the exlA gene comprises a single, 49 bp intron (position 581-629, see Fig. 1).

1.7 Determination of the 3' end of the exlA gene

The position of the stop codon of the exlA gene (position 1033-1035 in Fig. 1) was derived from DNA sequence data. This stop codon was confirmed, since the amino acid sequence of one of the peptides derived from endoxylanase II by chemical cleavage with CNBr proved to be identical to the C-terminal amino acid sequence derived from DNA sequence data (position 991-1032 in Fig. 1).

1.8 Evaluation of DNA and protein data

On the basis of the above data it was established that the gene coding for endoxylanase II of *Aspergillus niger* var. *awamori* had been cloned on a 5.3 kb *Sal*I fragment. The DNA sequence of the gene, the position of the intron and the length of the MRNA were established. The established N-terminal amino acid sequence of the mature protein was fully confirmed by the DNA sequence. On the basis of the above data it can be concluded that the exlA gene codes for a protein of 211 amino acids and that the first 27 amino acids are removed post-translationally. From this data the exlA signal sequence was derived.

Also, the nucleotide sequence of the exlA promoter follows from the obtained sequence (see Fig. 1, position 1-350). Also, the nucleotide sequence of the exlA

terminator follows from the obtained sequence (see Fig. 1, position 1036-2059 or a part thereof).

Example 2 Expression of the *Escherichia coli* ß-glucuronidase (uidA) gene using the exlA promoter and terminator sequences.

2.1 Construction of the uidA expression vector

The uidA expression plasmid (pAW15-1) was constructed starting from plasmid pAW14B, which contains a ca. 5.3 kb *Sal*I fragment on which the 0.7 kb endoxylanase II (exlA) gene is located, together with 2.5 kb of 5'-flanking sequences and 2.0 kb of 3'-flanking sequences (Fig.3). In pAW14B the exlA coding region was replaced by the uidA coding region. A *Bsp*HI site (5'-TCATGA-3') comprising the first codon (ATG) of the exlA gene and an *Afl*II site (5'-CTTAAG-3'), comprising the stopcodon (TAA) of the exlA gene facilitated the construction of pAW15-1.

The construction was carried out as follows: pAW14B (7.9 kb) was cut partially with *Bsp*HI (pAW14B contains five *Bsp*HI sites) and the linearized plasmid (7.9 kb) was isolated from an agarose gel. Subsequently the isolated 7.9 kb fragment was cut with *Bsm*I, which cuts a few nucleotides downstream of the *Bsp*HI site of interest, to remove plasmids linearized at the other *Bsp*HI sites. The fragments were separated on an agarose gel and the 7.9 kb *Bsp*HI-*Bsm*I fragment was isolated. This was partially cut with *Afl*II and the resulting 7.2 kb *Bsp*HI-*Afl*II fragment was isolated. The uidA gene was isolated as a 1.9 kb *Nco*I - *Afl*II fragment from pNOM-*Afl*II, a plasmid derived from pNOM102 (Roberts et al., 1989). In pNOM102 two *Nco*I sites are present, one of which is located at the 5'-end of the uidA gene and comprises the ATG-startcodon for translation of the gene. The second *Nco*I site is located a few nucleotides downstream of the stopcodon. To obtain an *Afl*II site downstream of the uidA stopcodon the latter *Nco*I site was converted into an *Afl*II site: pNOM102 was cut partially with *Nco*I and ligated with a *Nco*I - *Afl*II linker (Nco-Afl, see Table E), resulting in vector pNOM-*Afl*II. The 7.2 kb *Bsp*HI - *Afl*II fragment of pAW14B was ligated to the 1.9 kb *Nco*I - *Afl*II fragment of pNOM-*Afl*II to give vector pAW15-1 (Fig. 4).

The constructed vector (pAW15-1) can subsequently be transferred to moulds (for example *Aspergillus niger*, *Aspergillus niger* var. *awamori*, *Aspergillus nidulans* etc.) by

means of conventional co-transformation techniques and the ß-glucuronidase can then be expressed via induction of the endoxylanase II promoter. The constructed vector can also be provided with conventional selection markers (e.g. amdS or pyrG, hygromycin etc.) and the mould can be transformed with the resulting vector to produce the desired protein. As an example, the *E. coli* hygromycin selection marker was introduced in the uidA expression vector, yielding pAW15-7 (Fig. 4). For this purpose a fragment containing the *E. coli* hygromycin resistance gene controlled by the *Aspergillus nidulans* gpdA promoter and the *Aspergillus nidulans* trpC terminator was used. This cassette was isolated as a 2.6 kb *Not*I fragment from pBluekan7-1 in which the hygromycin resistance cassette is flanked by *Not*I sites. In pAW15-1 a *Not*I site was created by converting the *Eco*RI site (present 1.2 kb upstream of the ATG codon) into a *Not*I site using a synthetic oligonucleotide (Eco-Not, see Table E), yielding pAW15-1-Not. The 2.6 kb *Not*I fragment from pAWBluekan7-1 was isolated and ligated with *Not*I-linearized pAW15-1-Not. The resulting vector was called pAW15-7 (Fig.5).

2.2 Production of *E. coli* ß-glucuronidase driven by exlA expression signals

pAW15-7 was used to transform *Aspergillus niger* var. *awamori*. Transformant AW15.7-1 was identified by hygromycin selection and by Southern hybridization analysis of genomic DNA of this transformant it was established that this transformant contains a single copy of the uidA gene.

Aspergillus niger var. *awamori* (AW) and transformant AW15.7-1 were grown under the following conditions: shake flasks (500 ml) with 200 ml synthetic media (pH 6.5) were inoculated with spores (final concentration: 10E6/ml).

The medium had the following composition (AW Medium):

sucrose	10	g/l	$NaNO_3$	6.0	g/l
KCl	0.52	g/l	KH_2PO_4	1.52	g/l
$MgSO_4 \cdot 7H_2O$	0.49	g/l	Yeast extract	1.0	g/l
$ZnSO_4 \cdot 7H_2O$	22	mg/l	H_3BO_3	11	mg/l
$MnCl_2 \cdot 4H_2O$	5	mg/l	$FeSO_4 \cdot 7H_2O$	5	mg/l
$CaCl_2 \cdot 6H_2O$	1.7	mg/l	$CuSO_4 \cdot 5H_2O$	1.6	mg/l
$NaH_2MoO_4 \cdot 2H_2O$	1.5	mg/l	Na_2EDTA	50	mg/l

Incubation took place at 30°C, 200 rpm for 24 hours in a Mk X incubator shaker. After growth cells were removed by filtration (0.45 µm filter), washed twice with AW Medium without sucrose and yeast extract (salt solution), resuspended in 50 ml salt solution and transferred to 300 ml shake flasks containing 50 ml salt solution to which xylose has been added to a final concentration of 10 g/l (induction medium). The moment of resuspension is referred to as "t=0" (start of induction). Incubation took place under the same conditions as described above. Samples were taken 15 and 22 hours after induction. Biomass was recovered by filtration over miracloth, dried by squeezing and immediately frozen in liquid nitrogen. The mycelium was disrupted by grinding the frozen mycelium and ß-glucuronidase activity was determined essentially as described in Roberts et al. (1989)

From Table C it is evident that the exlA promoter is specifically induced by the presence of xylose, and that the exlA promoter and terminator can be used for the production of *E. coli* ß-glucuronidase in transformant AW15.7-1.

Strain	exp.	t=0	t=15	t=22
AW	A	0.0	0.0	0.1
AW	B	0.0	0.1	0.0
AW15.7-1	A	0.7	1110	823
AW15.7-1	B	0.6	1065	773

Table C. ß-glucuronidase production

Transformants were grown on synthetic medium as indicated in the text for 24 hours and at t=0 were transferred to induction medium as indicated in the text. ß-glucuronidase activity in the mycelium was determined as described in the text and is expressed in arbitrary units of enzymatic activity per milligram total protein.

Example 3 Production and secretion of the *Thermomyces lanuginosa* lipase using the exlA promoter, signal sequence and terminator.

3.1 Construction of expression plasmids based on the exlA expression signals

3.1.1 Vector

Plasmid pAW14A-Not was the starting vector for construction of a series of expression plasmids containing the exlA expression signals and the gene coding for *Thermomyces lanuginosa* lipase. Plasmid pAW14A comprises an *Aspergillus niger* var. *awamori* chromosomal 5 kb *Sal*I fragment on which the 0.7 kb exlA gene is located, together with 2.5 kb of 5'-flanking sequences and 2.0 kb of 3'-flanking sequences (similar to pAW14B, see Fig. 3). In pAW14A the *Eco*RI site originating from the pUC19 polylinker was converted to a *Not*I site by insertion of a synthetic oligonucleotide (Eco-Not, see Table E), yielding pAW14A-Not.

Starting from pAW14A-Not, constructs were made in which the exlA promoter (2.5 kb) was fused to the translation-initiation codon (ATG) of the *Thermomyces lanuginosa* lipase gene. Also, constructs were made in which the exlA promoter and the DNA sequence coding for the first 27 amino acids of the exlA protein, which is the preprosequence, was fused to the sequences coding for the mature lipase polypeptide.

In both series of expression vectors the exlA transcription terminator was used.

The following vector fragments were isolated from pAW14A-Not and used for the constructions:

* for the fusion with the translation-initiation codon of the lipase a 7.2 kb *Bsp*HI-*Afl*II fragment was isolated from pAW14A-Not. This is a similar fragment as the one isolated for the construction of pAW15-1 (see example 2) and was isolated essentially by the same approach as described in Example 2. The fragment contains 2.5 kb nucleotide sequences comprising the exlA promoter up to the *Bsp*HI site which comprises the ATG codon, and 2.0 kb nucleotide sequences comprising the exlA transcription terminator starting with the *Afl*II site which comprises the stopcodon.
* for the fusion of the exlA promoter and exlA pre-pro-peptide encoding region (the first 27 amino acids of the exlA gene) with the coding region of the mature lipase polypeptide, a partially digested 7.2 kb *Nru*I-*Afl*II fragment was isolated

from pAW14A-Not. This fragment contains 2.5 kb nucleotide sequences comprising the exlA promoter and the coding sequence of the first 26 amino acids of the exlA protein (preprosequence), which ends with a *Nru*I site, thus lacking only 1 amino acid of the exlA prosequence. Furthermore the fragment comprises 2 kb nucleotide sequences comprising the exlA transcription terminator starting with the *Afl*II site.

The *Thermomyces lanuginosa* lipase gene was isolated from vector pTL-1, which comprises a 0.9 kb coding region of the lipase gene (Figure 9) flanked by the *Aspergillus niger* glaA promoter and the *Aspergillus nidulans* trpC transcription terminator (Figure 8).

3.1.2 Fusion of the lipase gene with the exlA transcription terminator sequence

To obtain a fusion of the exlA transcription terminator with the lipase gene, an *Afl*II site was created just downstream of the stopcodon of the lipase gene. In pTL-1 a *Hin*dIII site is present 5 base pairs downstream of the stopcodon of the lipase gene (Figures 8 and 9), in which an *Afl*II site was created using a synthetic oligonucleotide. The construction was carried out as follows:

pTL-1 was cut with *Hin*dIII, yielding a linear 8.3 kb fragment, which was isolated from an agarose gel and ligated with the oligonucleotide Hind-Afl (see Table E). In the resulting vector, pTL1-AflII the *Hin*dIII site has disappeared and an *Afl*II site has been created just downstream of the stopcodon of the lipase gene, thus preparing the lipase gene for fusion to the exlA terminator at the *Afl*II site.

3.1.3 Fusion of the exlA promoter with the lipase gene (ATG fusion)

pTL1-AflII was used as starting vector to isolate a DNA fragment comprising the lipase gene. To fuse the lipase gene to the exlA promoter the region of the lipase gene comprising the ATG codon (ATATGA) was converted to a *Bsp*HI site (TCATGA). This site still comprises the correct coding sequence of the lipase gene. For this purpose a synthetic DNA fragment was used, consisting of oligonucleotides BTFF09 and BTFF10 (see Table E) annealed to each other. This synthetic fragment contains a *Xho*I site for cloning, followed by a *Bsp*HI site comprising the ATG codon and the next 7 base pairs of the lipase gene up to the *Sac*I site.

Vector pTL1-*Afl*II was linearized by partial digestion with *Xho*I, followed by cutting with *Sac*I which cuts after position +10 of the open reading frame encoding the lipase pre-pro-polypeptide. The 6.3 kb *Xho*I-*Sac*I vector fragment (resulting from a cut at the *Xho*I site in the glaA promoter while leaving the internal *Xho*I site in the lipase gene intact, see Figure 8) was isolated from an agarose gel and ligated with the synthetic *Xho*I-*Sac*I fragment resulting in vector pTL1-XS. From pTL1-XS a 0.9 kb *Bsp*HI-*Afl*II fragment comprising the lipase gene was isolated and ligated to the 7.2 kb *Bsp*HI-*Afl*II fragment from pAW14A-Not yielding expression vector pAWTL-1 (Fig.6).

3.1.4 Fusion of the exlA promoter and the region encoding the exlA prepro sequence with the coding sequence of the lipase mature protein

pTL1-*Afl*II was used as starting vector to isolate a DNA fragment comprising the lipase gene. To obtain a correct fusion of the sequence encoding the lipase mature polypeptide with the exlA promoter sequence and the exlA leader peptide encoding sequences, a synthetic DNA fragment was used, consisting of oligonucleotides BTFF05 and BTFF06 (see Table E) annealed to each other. This synthetic fragment comprises sequences encoding the last amino acid of the exlA pre-pro-sequence fused to the first 12 codons of the mature lipase encoding sequence. It contains a *Xho*I site for cloning and a *Nru*I site, which comprises the last 3 base pairs of the exlA prosequence. The fragment ends with a *Bgl*II site. Vector pTL1-*Afl*II was linearized by partial digestion with *Xho*I, followed by cutting with *Bgl*II, which cuts just within the region coding for the mature lipase. The 6.3 kb *Xho*I-*Bgl*II vector fragment (resulting from a cut at the *Xho*I site in the glaA promoter while leaving the internal *Xho*I site in the lipase gene intact, see Figure 8) was isolated from an agarose gel and ligated with the synthetic *Xho*I-*Bgl*II fragment, resulting in pTL1-XB. From pTL1-XB an 0.83 kb *Nru*I-*Afl*II fragment was isolated containing the last 3 base pairs of the exlA prosequence followed by the sequence encoding the mature lipase polypeptide up to the *Afl*II site just beyond the stop codon (see example 2). This fragment was ligated with the 7.2 kb *Nru*I-*Afl*II fragment of pAW14A-Not to give expression vector pAWTL2 (Fig.7).

3.2 Production and secretion of *Thermomyces lanuginosa* lipase using the exlA promoter and terminator.

The constructed expression vectors (pAWTL1 and pAWTL2) can subsequently be transferred to moulds (for example *Aspergillus niger*, *Aspergillus niger* var. *awamori*, *Aspergillus nidulans* etc.) by means of conventional co-transformation techniques and the lipase can then be expressed via induction of the endoxylanase II promoter. The constructed vector can also be provided with conventional selection markers (e.g. amdS or pyrG, hygromycin etc.) and the mould can be transformed with the resulting vector to produce the desired protein, essentially as described in example 2. As an example, plasmids were derived from pAWTL1 and pAWTL2 by introduction of an *Aspergillus niger* var. *awamori* pyrG gene, and the resulting plasmids were introduced in strain AWPYR, an *Aspergillus niger* var. *awamori* strain derived from strain CBS 115.52 (ATCC 11358) in which the pyrG gene has been disrupted. Following this route, transformant AWLPL1-2 was derived using the pAWTL1 plasmid, whereas transformant AWLPL2-2 was derived starting from the pAWTL2 plasmid.

Transformant AWLPL1-2 (containing the *Thermomyces lanuginosa* mature lipase encoding region with the *Thermomyces lanuginosa* signal sequence under the control of *Aspergillus niger* var. *awamori* exlA promoter and terminator) and transformant AWLPL2-2 (containing the *Thermomyces lanuginosa* mature lipase encoding region with the endoxylanase signal sequence under the control the *Aspergillus niger* awamori exlA promoter and terminator) were grown in shake flasks on AW Medium as described in example 2. Incubation took place at 30°C, 200 rpm for 24 hours in a Mk X incubator shaker. After growth cells were collected by filtration (0.45 µm filter), washed twice with AW Medium without sucrose and yeast extract (salt solution), resuspended in 50 ml salt solution and transferred to 300 ml shake flasks containing 50 ml salt solution to which xylose has been added to a final concentration of 10 g/l (induction medium). The moment of resuspension is referred to as "t=0" (start of induction). Incubation took place under the same conditions as described above. Samples were taken 15, 22 and 39 hours after induction. Samples were filtered over miracloth to remove biomass and the filtrate was analyzed for lipase activity by a titrimetric assay using olive oil as a substrate.

For each sample between 100 and 200 µl of filtrate was added to a stirred mixture of 5.0 ml lipase substrate (Sigma, containing olive oil as a substrate for the lipase) and 25.0 ml of buffer (5 mM Tris-HCl pH 9.0, 40 mM NaCl, 20 mM $CaCl_2$). The assay was carried out at 30°C and the release of fatty acids was measured by automated titration with 0.05 M NaOH to pH 9.0 using a Mettler DL25 titrator. A curve of the amount of titrant against time was obtained. The amount of lipase activity contained in the sample was calculated from the maximum slope of this curve. One unit of enzymatic activity is defined as the amount of enzyme that releases 1 µmol of fatty acid from olive oil in one minute under the conditions specified above. Such determinations are known to those skilled in the art.

The results are presented in Table D. From these results it is obvious that functional lipase is produced and secreted after induction of the exlA promoter by xylose, and that the exlA signal sequence can direct the secretion of heterologous proteins in *Aspergillus niger* var. *awamori*.

Strain	exp	t=0	t=15	t=22	t=39
AWLPL1-2	A	3.2	76	65	59
AWLPL1-2	B	7	84	32	35
AWLPL2-2	A	9	77	50	49
AWLPL2-2	B	8	72	51	46
AW	A	7	9	8	8
AW	B	6	7	7	8

Table D. Production and secretion of lipase

Transformants were grown on synthetic medium as indicated in the text for 24 hours and at t=0 were transferred to induction medium as indicated in the text. Lipase activity in the medium was determined by a titrimetric assay using olive oil as substrate and is expressed in arbitrary units of lipase activity. A and B represent duplo experiments.

BTFF05	5'-TCGAGTCGCGAGAGGTCTCGCAA-3'	sequence listing 5
BTFF06	5'-GATCTTGCGAGACCTCTCGCGAC-3'	sequence listing 6
BTFF09	5'-TCGAGCGTCATGAGGAGCT-3'	sequence listing 7
BTFF10	5'-CCTCATGACGC-3'	sequence listing 8
Eco-Not	5'-AATTGCGGCCGC-3'	sequence listing 9
Hind-Afl	5'-AGCTCGCTTAAGCG-3'	sequence listing 10
Nco-Afl	5'-CATGCCTTAAGG-3'	sequence listing 11
Xyl11	5'-GCATATGATTAAGCTGC-3'	sequence listing 12

Table E. Nucleotide sequences of oligonucleotides used in constructions

Sequence listing numbers refer to the listings provided in the official format.

Literature References

- Fuller, R.S., Sterne, R.E. and Thorner J. (1988) Enzymes required for yeast prohormone processing. Ann. Rev. Physiol. 50, 345-362
- van Heijne, G. (1986) A new method for preceding signal sequence cleavage sites. Nucl. Acids Res. 16, 4683-4690.
- Roberts, N., Oliver, R.P., Punt, P.J. and van den Hondel, C.A.M.J.J. (1989) Expression of the Escherichia coli β-glucuronidase gene in industrial and phytopathogenic filamentous fungi. Curr. Genet. 15, 177-180.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A laboratory manual (2nd ed). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sanger, F., Nicklan, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Ward, M., Wilson, L.J., Kodama, K.H., Rey, M.W. and Berka, R.M. (GENENCOR) (May 1990) Improved Production of Chymosin in *Aspergillus* by expression as a glucoamylase-chymosin fusion. Bio/Technology 8, 435-440.
- CA-A-2024448 (ALLELIX BIOPHARMACE) "Recombinant DNA expression construct - containing promoter for use in *Aspergillus*", published on 1 March 1991

- EP-A-0436858 (GREEN CROSS CORP.) "Promoter of glyceraldehyde-3-phosphate dehydrogenase gene - derived from *Aspergillus orizae*, used in new expression system in yellow-green or black koji mould", published on 17 July 1991
- EP-A-0439997 (CIBA GEIGI AG) "*A. niger* pyruvate kinase promoter - used to construct vectors for expression of structural genes in suitable hosts", published on 7 August 1991
- WO 91/19782 (UNILEVER) "Xylanase production", published on 26 December 1991, thus within the priority year

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Unilever N.V.
(B) STREET: Weena 455
(C) CITY: Rotterdam
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): NL-3013 AL

(A) NAME: Unilever PLC
(B) STREET: Unilever House Blackfriars
(C) CITY: London
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): EC4P 4BQ

(ii) TITLE OF INVENTION: Process for the production of a protein using endoxylanase II (exlA) expression signals

(iii) NUMBER OF SEQUENCES: 12

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2059 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Aspergillus niger var. awamori
(B) STRAIN: CBS 115.52 (ATCC 11358)

(vii) IMMEDIATE SOURCE:
(B) CLONE: pAW14B

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 581..629
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..350
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 351..431

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: join(432..580, 630..1032)
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /EC_number= 3.2.1.8
/product= "endoxylanase II"
/evidence= EXPERIMENTAL
/gene= "exlA"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: join(351..580, 630..1035)
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /EC_number= 3.2.1.8
/product= "pre-pro endoxylanase II"
/evidence= EXPERIMENTAL
/gene= "exlA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GCCCTTTTA TCCGTCTGCC GTCCATTTAG CCAAATGTAG TCCATTTAGC CAAGTGCGGT      60

CATTTAGCC AAGACCAGTG GCTAGATTGG TGGCTACACA GCAAACGCAT GACTGAGACA     120

AACTATAGG ACTGTCTCTG GAAATAGGCT CGAGGTTGTT CAAGCGTTTA AGGTGATGCG     180

CAAAATGCA TATGACTAAG CTGCTTCATC TTGCAGGGGG AAGGGATAAA TAGTCTTTTT     240

GCAGAATAT AAATAGAGGT AGAGTGGGCT CGCAGCAATA TTGACCAGCA CAGTGCTTCT     300

TTCCAGTTG CATAAATCCA TTCACCAGCA TTTAGCTTTC TTCAATCATC ATG AAG         356
                                                        Met Lys
                                                        -27

TC ACT GCG GCT TTT GCA GGT CTT TTG GTC ACG GCA TTC GCC GCT CCT        404
al Thr Ala Ala Phe Ala Gly Leu Leu Val Thr Ala Phe Ala Ala Pro
25                  -20                 -15                 -10

TG CCG GAA CCT GTT CTG GTG TCG CGA AGT GCT GGT ATT AAC TAC GTG        452
al Pro Glu Pro Val Leu Val Ser Arg Ser Ala Gly Ile Asn Tyr Val
                -5                  1                   5

AA AAC TAC AAC GGC AAC CTT GGT GAT TTC ACC TAT GAC GAG AGT GCC        500
ln Asn Tyr Asn Gly Asn Leu Gly Asp Phe Thr Tyr Asp Glu Ser Ala
       10                  15                  20
```

Tyr Gly Asp Tyr Asn Pro Cys Ser Ser Ala Thr Ser Leu Gly Thr Val
90 95 100

Tyr Ser Asp Gly Ser Thr Tyr Gln Val Cys Thr Asp Thr Arg Thr Asn
105 110 115

Glu Pro Ser Ile Thr Gly Thr Ser Thr Phe Thr Gln Tyr Phe Ser Val
120 125 130

Arg Glu Ser Thr Arg Thr Ser Gly Thr Val Thr Val Ala Asn His Phe
135 140 145

Asn Phe Trp Ala Gln His Gly Phe Gly Asn Ser Asp Phe Asn Tyr Gln
150 155 160 165

Val Met Ala Val Glu Ala Trp Ser Gly Ala Gly Ser Ala Ser Val Thr
170 175 180

Ile Ser Ser

TCCAGATATT CTATACTAAC AGACTTCTAA TGACTGCGGA TAATATAGAG GGCAAGAATT 1442

TCTACAGTTC GACGCAGTTC AACGCAATCA GAGAGGGAAT ACTGATGAGA GTGCAATCAG 1502

TTAGAGAAGG ACAACATGGC AGTCTTAGTG TGAACTTACA TAACGATATG GACTCTAGAA 1562

AAAAGGAAGG AGCTCCGTCT ATATATAGCG CCATTACGTG TATCTGATGC TTGCCCATTG 1622

CCACTGGGTA GGGTGACTTT TTGAAGCGAC TCGACATATA ATATGACAAA CTCATGCCCC 1682

CTTTGCAGGA AACTTAGCTT TTCCTGCCTT GCTTTGAAGC CACAATTATC ACGAAACTCA 1742

TTTAGAGATT TATCTTCCTG TAACGGAAAC AAATATTTCG GGATTGGAAT AGCCTTTTGC 1802

CGAACTCATT ATTTTTTTGC GACGGTAAAT CTGGGAGTAT ACGATGTCCT TTCACGTTTC 1862

TCAACAAAAC TCTGCCGCAC CGGGTAACCT ACGGATAGTA CTGTATCCAG ACTCAGTTTT 1922

TCTAATAACA GGACACTGTG CAATTTGCGG GAAAATTCCT ATGTATATTA CTTTCTCGTT 1982

GCATCTCAAA TATTGTGGCT TTTTGAGACC CACACTATGT CTTGCACATA TTGTACCATC 2042

CTTGCTTGAG GCCAATT 2059

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 211 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Val Thr Ala Ala Phe Ala Gly Leu Leu Val Thr Ala Phe Ala
-27     -25                 -20                 -15

Ala Pro Val Pro Glu Pro Val Leu Val Ser Arg Ser Ala Gly Ile Asn
    -10                 -5                  1                   5

Tyr Val Gln Asn Tyr Asn Gly Asn Leu Gly Asp Phe Thr Tyr Asp Glu
                10                  15                  20

Ser Ala Gly Thr Phe Ser Met Tyr Trp Glu Asp Gly Val Ser Ser Asp
            25                  30                  35

Phe Val Val Gly Leu Gly Trp Thr Thr Gly Ser Ser Asn Ala Ile Thr
        40                  45                  50

Tyr Ser Ala Glu Tyr Ser Ala Ser Gly Ser Ser Ser Tyr Leu Ala Val
    55                  60                  65

Tyr Gly Trp Val Asn Tyr Pro Gln Ala Glu Tyr Tyr Ile Val Glu Asp
70                  75                  80                  85
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 886 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..876
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /product= "Thermomyces lanuginosa pre-pro lipase"
/evidence= EXPERIMENTAL

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 67..873
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /product= "Thermomyces lanuginosa lipase"
/evidence= EXPERIMENTAL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG AGG AGC TCC CTT GTG CTG TTC TTT GTC TCT GCG TGG ACG GCC TTG        48
Met Arg Ser Ser Leu Val Leu Phe Phe Val Ser Ala Trp Thr Ala Leu
-22     -20                 -15                 -10

GCC AGT CCT ATT CGT CGA GAG GTC TCG CAA GAT CTG TTT AAC CAG TTC        96
Ala Ser Pro Ile Arg Arg Glu Val Ser Gln Asp Leu Phe Asn Gln Phe
 -5                  1                   5                  10

AAT CTC TTT GCA CAG TAT TCT GCT GCC GCA TAC TGC GGA AAA AAC AAT       144
Asn Leu Phe Ala Gln Tyr Ser Ala Ala Ala Tyr Cys Gly Lys Asn Asn
                15                  20                  25

GAT GCC CCA GCT GGT ACA AAC ATT ACG TGC ACG GGA AAT GCC TGC CCC       192
Asp Ala Pro Ala Gly Thr Asn Ile Thr Cys Thr Gly Asn Ala Cys Pro
            30                  35                  40

GAG GTA GAG AAG GCG GAT GCA ACG TTT CTC TAC TCG TTT GAA GAC TCT       240
Glu Val Glu Lys Ala Asp Ala Thr Phe Leu Tyr Ser Phe Glu Asp Ser
        45                  50                  55

GGA GTG GGC GAT GTC ACC GGC TTC CTT GCT CTA GAC AAC ACG AAC AAA       288
Gly Val Gly Asp Val Thr Gly Phe Leu Ala Leu Asp Asn Thr Asn Lys
    60                  65                  70

TTG ATC GTC CTC TCT TTC CGT GGC TCT CGT TCC ATA GAA AAC TGG ATC       336
Leu Ile Val Leu Ser Phe Arg Gly Ser Arg Ser Ile Glu Asn Trp Ile
75                  80                  85                  90
```

```
GGA AAT CTT AAC TTC GAC TTG AAA GAA ATA AAT GAC ATT TGC TCC GGC          384
Gly Asn Leu Asn Phe Asp Leu Lys Glu Ile Asn Asp Ile Cys Ser Gly
                95                  100                 105

TGC AGG GGA CAT GAC GGC TTC ACC TCG AGC TGG AGG TCT GTA GCC GAT          432
Cys Arg Gly His Asp Gly Phe Thr Ser Ser Trp Arg Ser Val Ala Asp
            110                 115                 120

ACG TTA AGG CAG AAG GTG GAG GAT GCT GTG AGG GAG CAT CCC GAC TAT          480
Thr Leu Arg Gln Lys Val Glu Asp Ala Val Arg Glu His Pro Asp Tyr
        125                 130                 135

CGC GTG GTG TTT ACC GGA CAT AGC TTG GGT GGT GCA TTG GCA ACT GTT          528
Arg Val Val Phe Thr Gly His Ser Leu Gly Gly Ala Leu Ala Thr Val
    140                 145                 150

GCC GGA GCA GAC CTG CGT GGA AAT GGG TAT GAC ATC GAC GTG TTT TCA          576
Ala Gly Ala Asp Leu Arg Gly Asn Gly Tyr Asp Ile Asp Val Phe Ser
155                 160                 165                 170

TAT GGC GCC CCC CGA GTC GGA AAC AGG GCT TTT GCA GAA TTC CTG ACC          624
Tyr Gly Ala Pro Arg Val Gly Asn Arg Ala Phe Ala Glu Phe Leu Thr
                175                 180                 185

GTA CAG ACC GGC GGT ACC CTC TAC CGC ATT ACC CAC ACC AAT GAT ATT          672
Val Gln Thr Gly Gly Thr Leu Tyr Arg Ile Thr His Thr Asn Asp Ile
            190                 195                 200

GTC CCT AGA CTC CCG CCG CGC GAG TTC GGT TAC AGC CAT TCT AGC CCA          720
Val Pro Arg Leu Pro Pro Arg Glu Phe Gly Tyr Ser His Ser Ser Pro
        205                 210                 215

GAG TAC TGG ATC AAA TCT GGA ACC CTT GTC CCC GTC ACC CGA AAC GAC          768
Glu Tyr Trp Ile Lys Ser Gly Thr Leu Val Pro Val Thr Arg Asn Asp
    220                 225                 230

ATC GTG AAG ATA GAA GGC ATC GAT GCC ACC GGC GGC AAT AAC CAG CCT          816
Ile Val Lys Ile Glu Gly Ile Asp Ala Thr Gly Gly Asn Asn Gln Pro
235                 240                 245                 250

AAC ATT CCG GAT ATC CCT GCG CAC CTA TGG TAC TTC GGG TTA ATT GGG          864
Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly
                255                 260                 265

ACA TGT CTT TAGTGCGAAG CTT                                              886
Thr Cys Leu
          270
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 291 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Ser Ser Leu Val Leu Phe Phe Val Ser Ala Trp Thr Ala Leu
-22 -20 -15 -10

Ala Ser Pro Ile Arg Arg Glu Val Ser Gln Asp Leu Phe Asn Gln Phe
-5 1 5 10

Asn Leu Phe Ala Gln Tyr Ser Ala Ala Ala Tyr Cys Gly Lys Asn Asn
15 20 25

Asp Ala Pro Ala Gly Thr Asn Ile Thr Cys Thr Gly Asn Ala Cys Pro
30 35 40

Glu Val Glu Lys Ala Asp Ala Thr Phe Leu Tyr Ser Phe Glu Asp Ser
45 50 55

Gly Val Gly Asp Val Thr Gly Phe Leu Ala Leu Asp Asn Thr Asn Lys
60 65 70

Leu Ile Val Leu Ser Phe Arg Gly Ser Arg Ser Ile Glu Asn Trp Ile
75 80 85 90

Gly Asn Leu Asn Phe Asp Leu Lys Glu Ile Asn Asp Ile Cys Ser Gly
95 100 105

Cys Arg Gly His Asp Gly Phe Thr Ser Ser Trp Arg Ser Val Ala Asp
110 115 120

Thr Leu Arg Gln Lys Val Glu Asp Ala Val Arg Glu His Pro Asp Tyr
125 130 135

Arg Val Val Phe Thr Gly His Ser Leu Gly Gly Ala Leu Ala Thr Val
140 145 150

Ala Gly Ala Asp Leu Arg Gly Asn Gly Tyr Asp Ile Asp Val Phe Ser
155 160 165 170

Tyr Gly Ala Pro Arg Val Gly Asn Arg Ala Phe Ala Glu Phe Leu Thr
175 180 185

Val Gln Thr Gly Gly Thr Leu Tyr Arg Ile Thr His Thr Asn Asp Ile
190 195 200

Val Pro Arg Leu Pro Pro Arg Glu Phe Gly Tyr Ser His Ser Ser Pro
205 210 215

Glu Tyr Trp Ile Lys Ser Gly Thr Leu Val Pro Val Thr Arg Asn Asp
220 225 230

Ile Val Lys Ile Glu Gly Ile Asp Ala Thr Gly Gly Asn Asn Gln Pro
235 240 245 250

Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly
255 260 265

Thr Cys Leu

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCGAGTCGCG AGAGGTCTCG CAA 23

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GATCTTGCGA GACCTCTCGC GAC 23

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCGAGCGTCA TGAGGAGCT 19

CLAIMS

1. Process for the production and optionally secretion of a protein by means of a transformed mould, into which an expression vector has been introduced with the aid of recombinant DNA techniques known per se, said vector comprising one or more mould-derived expression and/or secretion regulating regions,
in which process at least one of said expression and/or secretion regulating regions is selected from

(1) the expression and secretion regulating regions of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and

(2) functional derivatives thereof also having expression and/or secretion regulating activity,

with the proviso that said protein is different from the endoxylanase type II ex *Aspergillus niger* var. *awamori*.

2. Process according to Claim 1, in which process the selected expression regulating region is a promoter and said vector comprises a gene encoding said protein under control of said promoter, the latter being selected from

(1) the promoter of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and

(2) functional derivatives thereof also having promoter activity.

3. Process according to Claim 2, in which said promoter is equal to the promoter present on the 5' part upstream of the exlA gene having a size of about 2.5 kb located between the *Sal*I restriction site at position 0 and the start codon ATG of the exlA gene in plasmid pAW14B.

4. Process according to Claim 2, in which said promoter comprises at least the polynucleotide sequence 1-350 according to Figure 1.

5. Process according to Claim 2, in which said promoter is induced by wheat bran, xylan, or xylose, or a mixture of any combination thereof, present in a medium in which the transformed mould is incubated.

6. Process according to claim 5, in which said promoter is induced by xylose present in the medium in which the transformed mould is incubated.

7. Process according to Claim 1, in which process the selected expression regulating region is a terminator and said vector comprises a gene encoding said protein followed by said terminator, the latter being selected from

(1) the terminator of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and

(2) functional derivatives thereof also having terminator activity.

8. Process according to Claim 7, in which said terminator is equal to the terminator present on the 3' part downstream of the exlA gene having a size of about 1.0 kb located right downstream of the stop codon (TAA) of the exlA gene in plasmid pAW14B.

9. Process according to Claim 7, in which said vector also comprises a promoter as claimed in Claim 2.

10. Process according to Claim 1 for the production and secretion of a protein by means of a transformed mould, in which process the selected secretion regulating region is a DNA sequence encoding a signal sequence and said vector comprises a gene encoding said protein preceded by said DNA sequence encoding a signal sequence, the latter being selected from

(1) the DNA sequence encoding the signal sequence of the endoxylanase II gene (exlA gene) of *Aspergillus niger* var. *awamori* present on plasmid pAW14B (Figure 3), which is present in a transformed *E. coli* strain JM109 deposited at the Centraalbureau voor Schimmelcultures in Baarn, The Netherlands, under N° CBS 237.90 on 31 May 1990, and

(2) functional derivatives thereof also directing secretion of the protein.

11. Process according to Claim 10, in which the gene (1) encoding said protein is also preceded by at least an essential part of a DNA sequence (2) encoding the mature endoxylanase II protein, whereby said DNA sequence (2) is present between said DNA sequence encoding a signal sequence (3) and the gene (1).

12. Process according to Claim 10. in which said signal sequence is the signal sequence encoded by polynucleotide 351-431 of the DNA sequence given in Figure 1, which polynucleotide precedes the exlA gene in plasmid pAW14B.

13. Process according to Claim 10, in which said vector also comprises a promoter as claimed in Claim 2 or a terminator as claimed in Claim 7, or both.

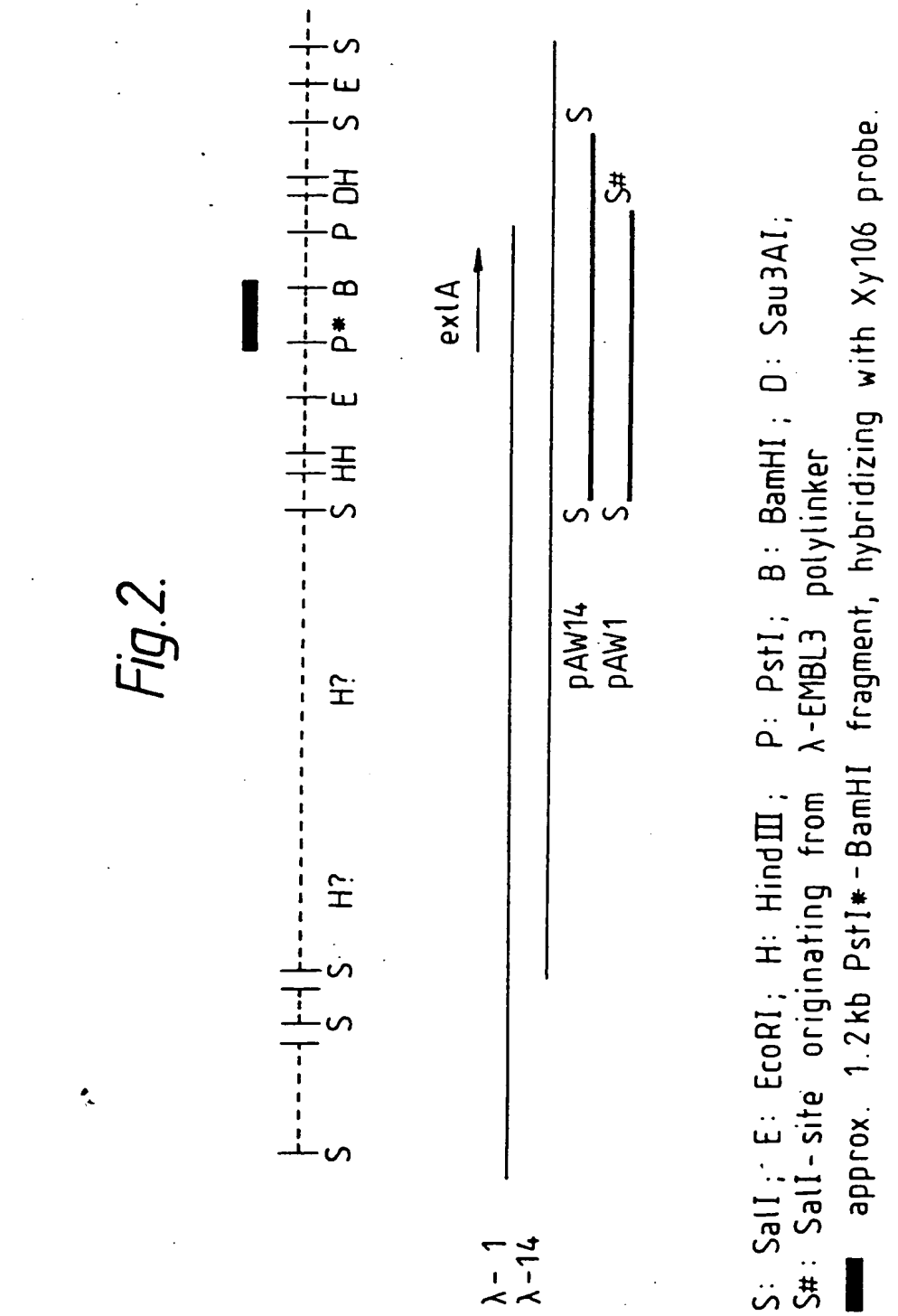
Fig. 1.

```
          10                  30                  50
AGcCCTTTTA TCCgTCTgcc gTCcATTTAG CCAAATGTAG TCCATTTAGC CAAGTGCGGT

          70                  90                 110
CCATTTAGCC AAGACCAGTG GCTAGATTGG TGGCTACACA GCAAACGCAT GACTGAGACA

         130                 150                 170
CAACTATAGG ACTGTCTCTG GAAATAGGCT CGAGGTTGTT CAAGCGTTTA AGGTGATGCG

         190                 210                 230
GCAAAATGCA TATGACTAAG CTGCTTcATC TTGCAGGGGG AAGGGATAAA TAGTCTTTTT

         250                 270                 290
CGCAGAATAT AAATAGAGGT AGAGTGGGCT CGCAGCAATA TTGACCAGCA CAGTGCTTCT

         310                 330                 350
CTTCCAGTTG CATAAATCCA TTCACCAGCA TTTAGCTTTC TTCAATCATC ATG AAG GTC
                                                       M   K   V

                        380                 400
ACT GCG GCT TTT GCA GGT CTT TTG GTC ACG GCA TTC GCC GCT CCT GTG CCG
T   A   A   F   A   G   L   L   V   T   A   F   A   A   P   V   P

          420                 440                 460
GAA CCT GTT CTG GTG TCG CGA AGT GCT GGT ATT AAC TAC GTG CAA AAC TAC
E   P   V   L   V   S   R   S   A   G   I   N   Y   V   Q   N   Y
                                  |> mature xylanase

                    480                 500
AAC GGC AAC CTT GGT GAT TTC ACC TAT GAC GAG AGT GCC GGA ACA TTT TCC
N   G   N   L   G   D   F   T   Y   D   E   S   A   G   T   F   S

        520                 540                 560
ATG TAC TGG GAA GAT GGA GTG AGC TCC GAC TTT GTC GTT GGT CTG GGC TGG
M   Y   W   E   D   G   V   S   S   D   F   V   V   G   L   G   W

                580                 600                 620
ACC ACT GGT TCT TCT AA GTGAGTGACT GTATTCTTTA ACCAAAGTCT AGGATCTAAC
T   T   G   S   S   N

                    640                 660
GTTTTCTAG C GCT ATC ACC TAC TCT GCC GAA TAC AGT GCT TCT GGC TCC TCT
            A   I   T   Y   S   A   E   Y   S   A   S   G   S   S

        680                 700                 720
TCC TAC CTC GCT GTG TAC GGC TGG GTC AAC TAT CCT CAG GCT GAA TAC TAC
S   Y   L   A   V   Y   G   W   V   N   Y   P   Q   A   E   Y   Y

                740                 760
ATC GTC GAG GAT TAC GGT GAT TAC AAC CCT TGC AGC TCG GCC ACA AGC CTT
I   V   E   D   Y   G   D   Y   N   P   C   S   S   A   T   S   L

    780                 800                 820
GGT ACC GTG TAC TCT GAT GGA AGC ACC TAC CAA GTC TGC ACC GAC ACT CGA
G   T   V   Y   S   D   G   S   T   Y   Q   V   C   T   D   T   R

                840                 860
ACT AAC GAA CCG TCC ATC ACG GGA ACA AGC ACG TTC ACG CAG TAC TTC TCC
T   N   E   P   S   I   T   G   T   S   T   F   T   Q   Y   F   S

  880                 900                 920
GTT CGA GAG AGC ACG CGC ACA TCT GGA ACG GTG ACT GTT GCC AAC CAT TTC
V   R   E   S   T   R   T   S   G   T   V   T   V   A   N   H   F
```

Fig. 1 (cont.)

```
                940                             960
AAC TTC TGG GCG CAG CAT GGG TTC GGA AAT AGC GAC TTC AAT TAT CAG GTC
N   F   W   A   Q   H   G   F   G   N   S   D   F   N   Y   Q   V

980                             1000                            1020
ATG GCA GTG GAA GCA TGG AGC GGT GCT GGC AGC GCC AGT GTC ACG ATC TCC
M   A   V   E   A   W   S   G   A   G   S   A   S   V   T   I   S

                        1050                    1070                    1090
TCT TAA GGGAT AAGTGCCTTG GTAGTCGGAA GATGTCAACG CGGAACTTTG TTCTCAGCTG
S   *

                    1110                    1130                    1150
GTGTGATGAT CGGATCCGGC CTCTGGTGGT TACATTGAGG CTGTATAAGT TATTCTGGGG

                    1170                    1190                    1210
CCGAGCTGTC AGCGGCTGCG TTTCCAATTT GCACAGATAA TCAACTTTCg TTTTCTATCT

                    1230                    1250                    1270
CTTGCGTTTC CACGCTGTTT ATCCTATCCA TAGATAATAT tTTgCCCAAT ACATATTATC

                    1290                    1310                    1330
TATATACAAC TTGTTCGGTC GCAGTAGTCA CTCCGAGCAA GGCATTGGGA AATTGGGAGA

                    1350                    1370                    1390
TGCGGGGTGC TGCGTACGCT CTAAGGTAGG GCATTTAAAG GGATATTTAG CCTCCAGATA

                    1410                    1430                    1450
TTCTATACTA ACAGACTTCT AATGACTGCG GATAATATAG AGGGCAAGAA TTTCTACAGT

                    1470                    1490                    1510
TCGACGCAGT TCAACGCAAT CAGAGAGGGA ATACTGATGA GAGTGCAATC AGTTAGAGAA

                    1530                    1550                    1570
GGACAACATG GCAGTCTTAG TGTGAACTTA CATAACGATA TGGACTCTAG AAAAAAGGAA

                    1590                    1610                    1630
GGAGCTCCGT CTATATATAG CGCCATTACG TGTATCTGAT GCTTGCCCAT TGCCACTGGG

                    1650                    1670                    1690
TAGGGTGACT TTTTGAAGCG ACTCGACATA TAATATGACA AACTCATGCC CCCTTTGCAG

                    1710                    1730                    1750
GAAACTTAGC TTTTCCTGCC TTGCTTTGAA GCCACAATTA TCACGAAACT CATTTAGAGA

                    1770                    1790                    1810
TTTATCTTCC TGTAACGGAA ACAAATATTT CGGGATTGGA ATAGCCTTTT GCCGAACTCA

                    1830                    1850                    1870
TTATTTTTTT GCGACGGTAA ATCTGGGAGT ATACGATGTC CTTTCACGTT TCTCAACAAA

                    1890                    1910                    1930
ACTCTGCCGC ACCGGGTAAC CTACGGATAG TACTGTATCC AGACTCAGTT TTTCTAATAA

                    1950                    1970                    1990
CAGGACACTG TGCAATTTGC GGGAAAATTC CTATGTATAT TACTTTCTCG TTGCATCTCA

                    2010                    2030                    2050
AATATTGTGG CTTTTTGAGA CCCACACTAT GTCTTGCACA TATTGTACCA TCCTTGCTTG

     2059
AGGCCAATT
```

Fig.2.



S: SalI; E: EcoRI; H: HindIII; P: PstI; B: BamHI; D: Sau3AI;
S#: SalI-site originating from λ-EMBL3 polylinker
▬ approx. 1.2kb PstI*-BamHI fragment, hybridizing with Xy106 probe.

Fig.3.

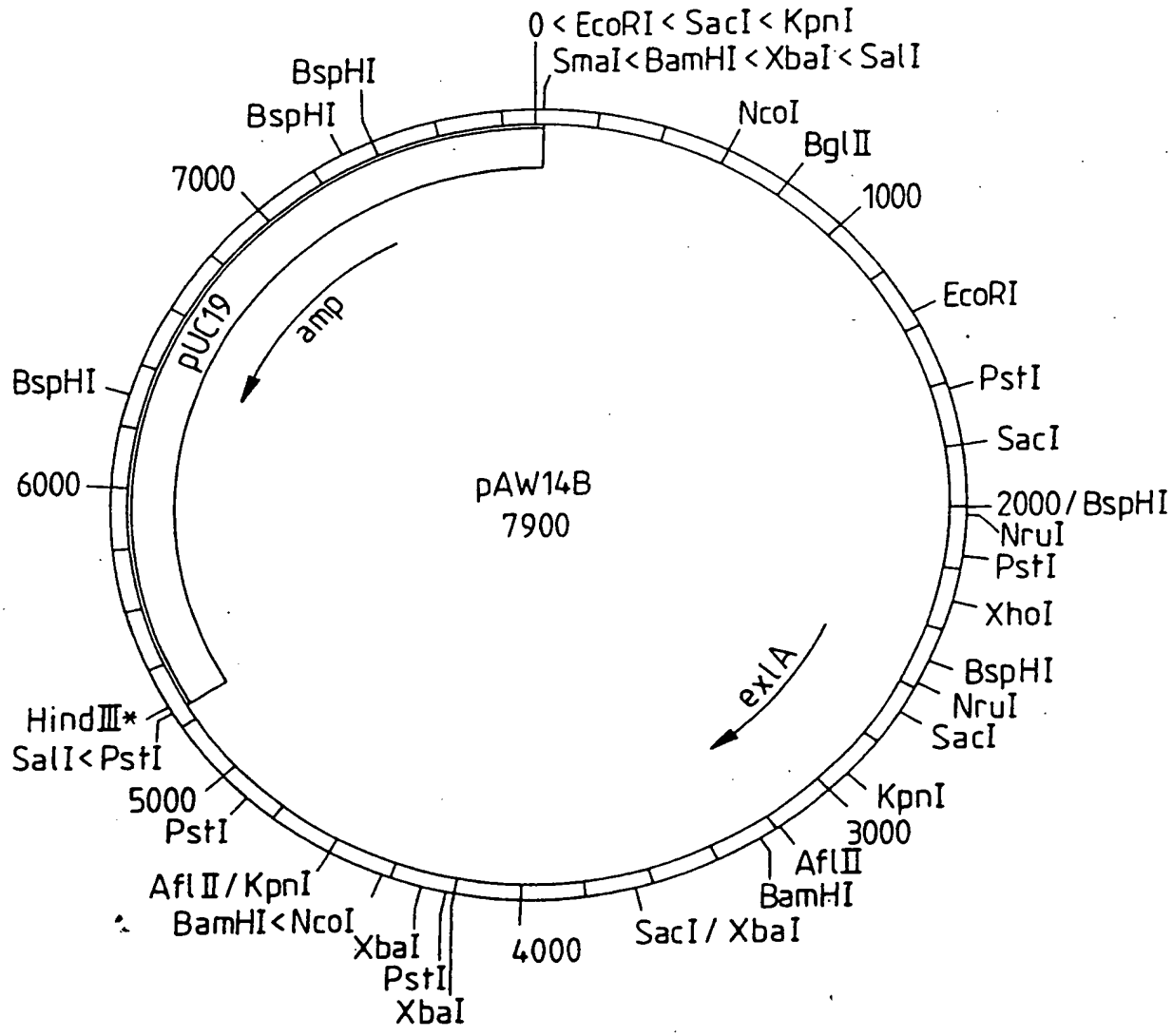

0 < EcoRI < SacI < KpnI
SmaI < BamHI < XbaI < SalI
NcoI
BglII
1000
EcoRI
PstI
SacI
2000 / BspHI
NruI
PstI
XhoI
BspHI
NruI
SacI
KpnI
3000
AflII
BamHI
SacI / XbaI
4000
XbaI
PstI
XbaI
AflII / KpnI
BamHI < NcoI
5000
PstI
HindIII*
SalI < PstI
6000
BspHI
7000
BspHI
BspHI
pUC19
amp
exlA
pAW14B
7900

Fig. 4.

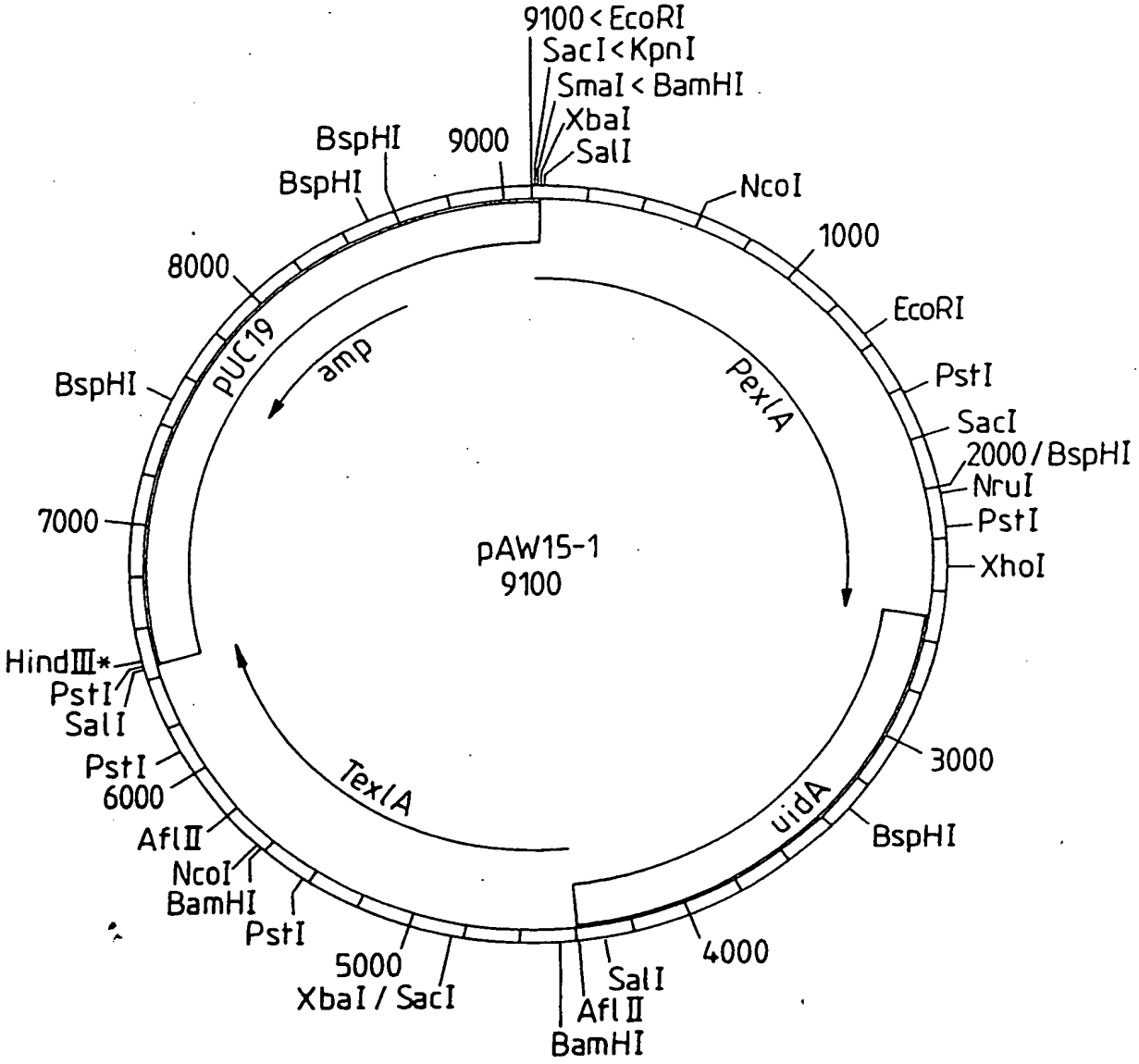

9100 < EcoRI
SacI < KpnI
SmaI < BamHI
XbaI
SalI
NcoI
1000
EcoRI
PstI
SacI
2000 / BspHI
NruI
PstI
XhoI
3000
BspHI
4000
SalI
AflII
BamHI
5000
XbaI / SacI
PstI
BamHI
NcoI
AflII
6000
PstI
SalI
PstI
HindIII*
7000
BspHI
8000
BspHI
BspHI
9000
pUC19
amp
pAW15-1
9100
PexIA
uidA
TexIA

Fig.5.

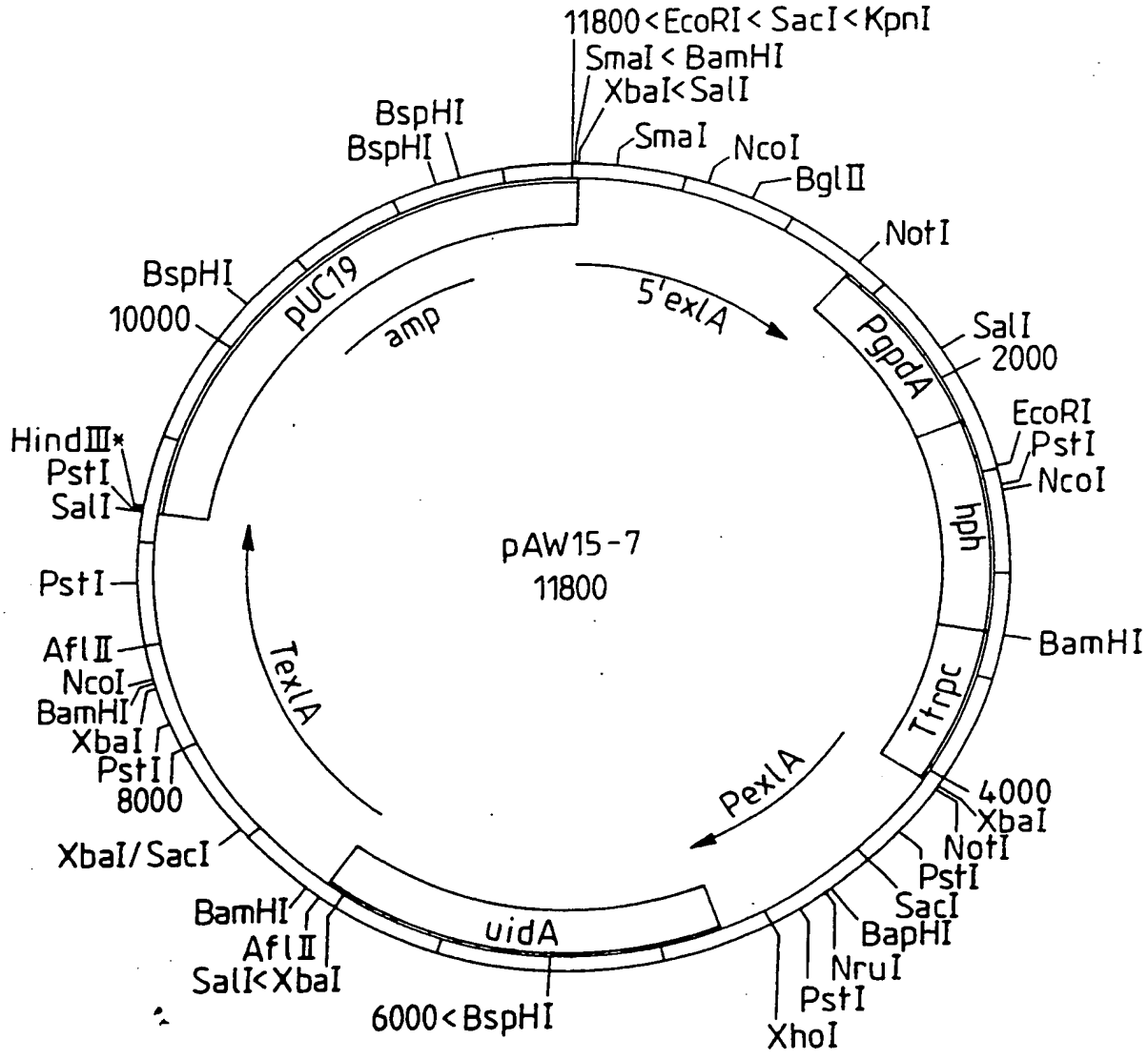

11800 < EcoRI < SacI < KpnI
SmaI < BamHI
XbaI < SalI
SmaI
NcoI
BglII
NotI
SalI
2000
EcoRI
PstI
NcoI
BamHI
4000
XbaI
NotI
PstI
SacI
BapHI
NruI
PstI
XhoI
6000 < BspHI
BamHI
AflII
SalI < XbaI
XbaI/SacI
8000
PstI
XbaI
BamHI
NcoI
AflII
PstI
SalI
HindIII*
BspHI
10000
BspHI
BspHI
pUC19
amp
5'exlA
PgpdA
hph
Trpc
PexlA
uidA
TexlA
pAW15-7
11800

Fig.6.

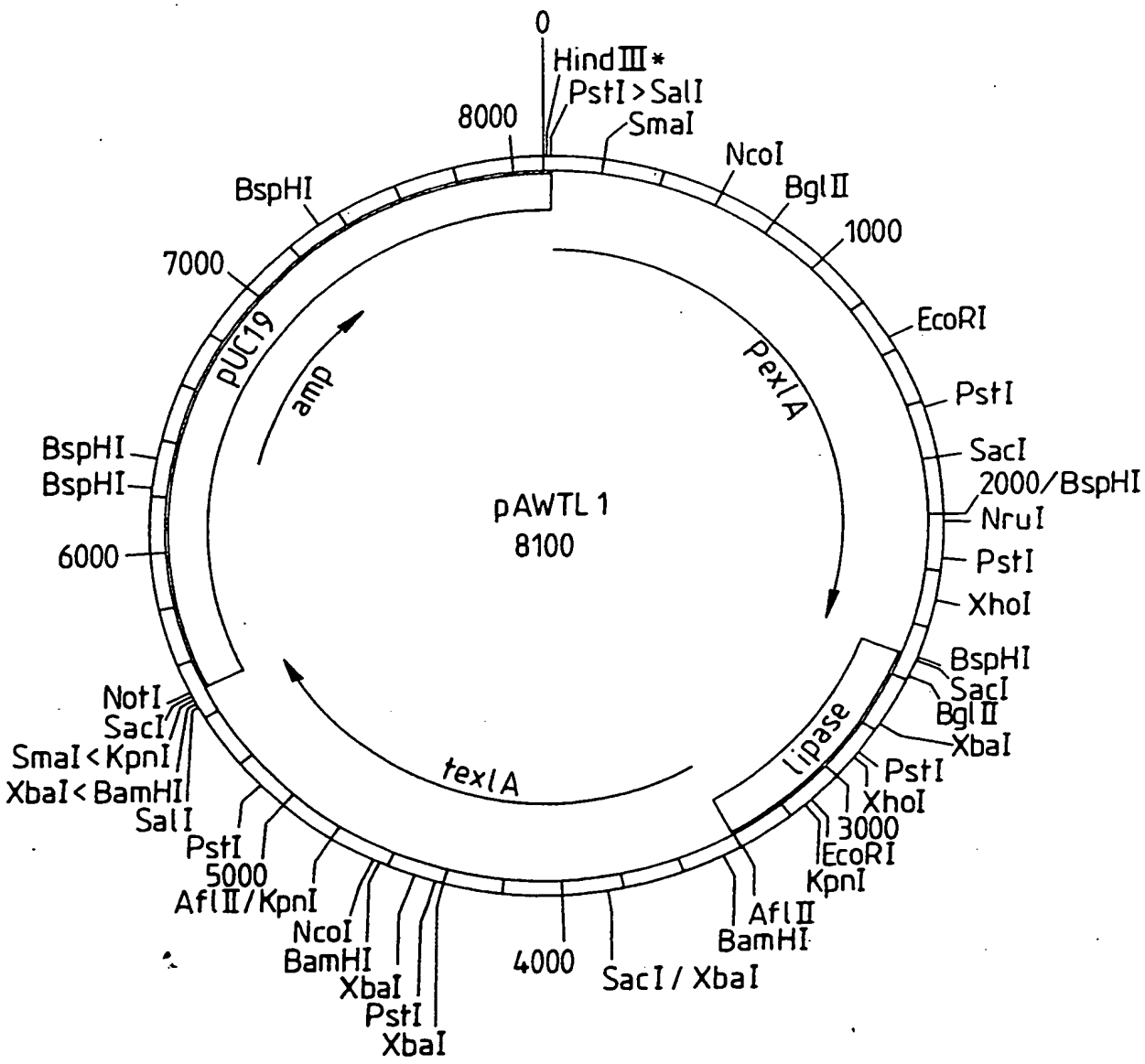

0
HindIII*
PstI > SalI
SmaI
NcoI
BglII
1000
EcoRI
PstI
SacI
2000/BspHI
NruI
PstI
XhoI
BspHI
SacI
BglII
XbaI
PstI
XhoI
3000
EcoRI
KpnI
AflII
BamHI
4000
SacI / XbaI
NcoI
BamHI
XbaI
PstI
XbaI
PstI
5000
AflII/KpnI
SalI
XbaI < BamHI
SmaI < KpnI
SacI
NotI
BspHI
BspHI
6000
BspHI
7000
8000
pUC19
amp
pAWTL 1
8100
texlA
pexlA
lipase

Fig. 7.

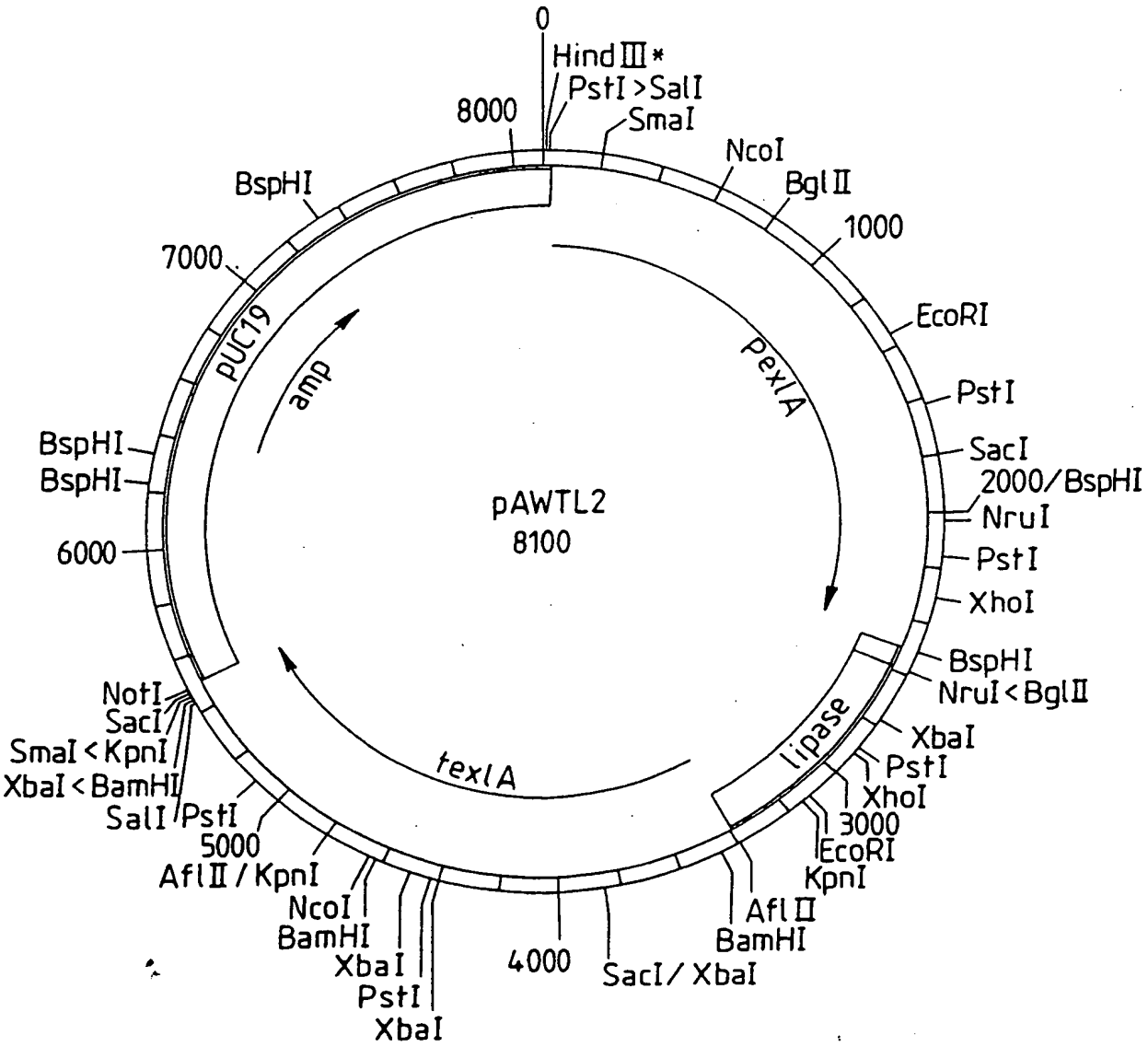

0
HindIII*
PstI>SalI
SmaI
NcoI
BglII
1000
EcoRI
PstI
SacI
2000/BspHI
NruI
PstI
XhoI
BspHI
NruI<BglII
XbaI
PstI
XhoI
3000
EcoRI
KpnI
AflII
BamHI
4000
SacI/XbaI
NcoI
BamHI
XbaI
PstI
XbaI
AflII/KpnI
5000
PstI
SalI
XbaI<BamHI
SmaI<KpnI
SacI
NotI
6000
BspHI
BspHI
7000
BspHI
8000
pAWTL2
8100
pUC19
amp
texlA
lipase
pexlA

Fig. 8.

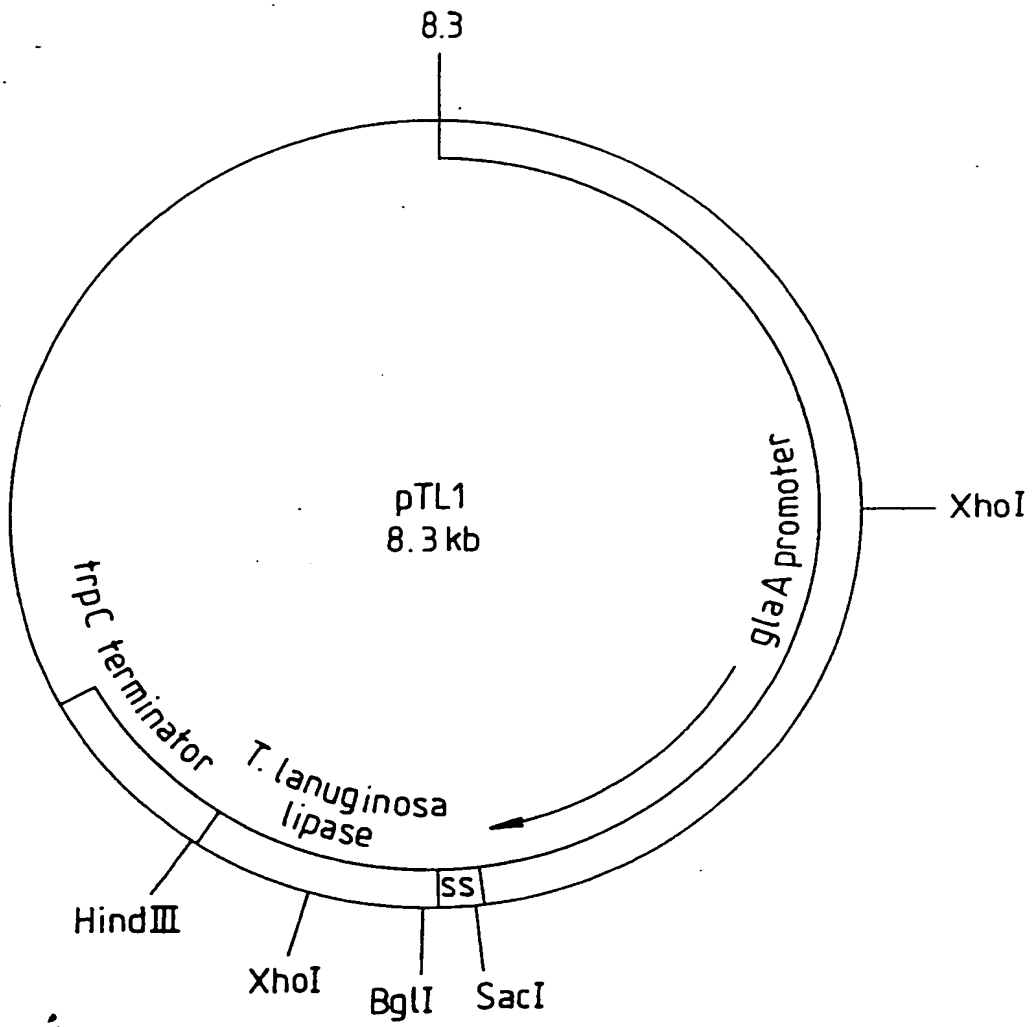

8.3
pTL1
8.3 kb
XhoI
glaA promoter
T. lanuginosa lipase
ss
trpC terminator
HindIII
XhoI
BglI
SacI

Fig. 9.

```
   1  ATG AGG AGC TCC CTT GTG CTG TTC TTT GTC TCT GCG TGG ACG GCC TTG
      M   R   S   S   L   V   L   F   F   V   S   A   W   T   A   L

  49  GCC AGT CCT ATT CGT CGA GAG GTC TCG CAA GAT CTG TTT AAC CAG TTC
      A   S   P   I   R   R   E   V   S   Q   D   L   F   N   Q   F
                              |> mature lipase

  97  AAT CTC TTT GCA CAG TAT TCT GCT GCC GCA TAC TGC GGA AAA AAC AAT
      N   L   F   A   Q   Y   S   A   A   A   Y   C   G   K   N   N

 145  GAT GCC CCA GCT GGT ACA AAC ATT ACG TGC ACG GGA AAT GCC TGC CCC
      D   A   P   A   G   T   N   I   T   C   T   G   N   A   C   P

 193  GAG GTA GAG AAG GCG GAT GCA ACG TTT CTC TAC TCG TTT GAA GAC TCT
      E   V   E   K   A   D   A   T   F   L   Y   S   F   E   D   S

 241  GGA GTG GGC GAT GTC ACC GGC TTC CTT GCT CTA GAC AAC ACG AAC AAA
      G   V   G   D   V   T   G   F   L   A   L   D   N   T   N   K

 289  TTG ATC GTC CTC TCT TTC CGT GGC TCT CGT TCC ATA GAA AAC TGG ATC
      L   I   V   L   S   F   R   G   S   R   S   I   E   N   W   I

 337  GGA AAT CTT AAC TTC GAC TTG AAA GAA ATA AAT GAC ATT TGC TCC GGC
      G   N   L   N   F   D   L   K   E   I   N   D   I   C   S   G

 385  TGC AGG GGA CAT GAC GGC TTC ACC TCG AGC TGG AGG TCT GTA GCC GAT
      C   R   G   H   D   G   F   T   S   S   W   R   S   V   A   D

 433  ACG TTA AGG CAG AAG GTG GAG GAT GCT GTG AGG GAG CAT CCC GAC TAT
      T   L   R   Q   K   V   E   D   A   V   R   E   H   P   D   Y

 431  CGC GTG GTG TTT ACC GGA CAT AGC TTG GGT GGT GCA TTC GCA ACT GTT
      R   V   V   F   T   G   H   S   L   G   G   A   L   A   T   V

 529  GCC GGA GCA GAC CTG CGT GGA AAT GGG TAT GAC ATC GAC GTG TTT TCA
      A   G   A   D   L   R   G   N   G   Y   D   I   D   V   F   S

 577  TAT GGC GCC CCC CGA GTC GGA AAC AGG GCT TTT GCA GAA TTC CTG ACC
      Y   G   A   P   R   V   G   N   R   A   F   A   E   F   L   T

 625  GTA CAG ACC GGC GGT ACC CTC TAC CGC ATT ACC CAC ACC AAT GAT ATT
      V   Q   T   G   G   T   L   Y   R   I   T   H   T   N   D   I

 673  GTC CCT AGA CTC CCG CCG CGC GAG TTC GGT TAC AGC CAT TCT AGC CCA
      V   P   R   L   P   P   R   E   F   G   Y   S   H   S   S   P

 721  GAG TAC TGG ATC AAA TCT GGA ACC CTT GTC CCC GTC ACC CGA AAC GAC
      E   Y   W   I   K   S   G   T   L   V   P   V   T   R   N   D

 769  ATC GTG AAG ATA GAA GGC ATC GAT GCC ACC GGC GGC AAT AAC CAG CCT
      I   V   K   I   E   G   I   D   A   T   G   G   N   N   Q   P

 817  AAC ATT CCG GAT ATC CCT GCG CAC CTA TGG TAC TTC GGG TTA ATT GGG
      N   I   P   D   I   P   A   H   L   W   Y   F   G   L   I   G

 865  ACA TGT CTT TAG TGCGAAGCTT 886
      T   C   L
```

INTERNATIONAL SEARCH REPORT

International Appli[redacted] No PCT/EP 92/02896

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/56; C12N15/62; C12N15/80; //C12N15/55

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
Y	BIO/TECHNOLOGY vol. 8, no. 5, May 1990, NATURE AMERICA, INC., NEW YORK, US pages 435 - 440 M. WARD ET AL. 'Improved production of chymosin in Aspergillus by expression as a glucoamylase-chymosin fusion' cited in the application see page 435, right column, line 56 - page 436, left column, line 7 see page 438, right column, line 4 - line 7 see page 439, right column, line 1 - line 13 ---	10-13
Y	EP,A,0 421 919 (CIBA-GEIGY AG) 10 April 1991 see page 16, line 3 - line 10 --- -/--	1-13

° Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
02 APRIL 1993	
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	HORNIG H.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	EP,A,0 357 127 (GIST-BROCADES N.V.) 7 March 1990 see page 4, line 3 - line 42	1-13
P,Y	EP,A,0 463 706 (GIST-BROCADES N.V.) 2 January 1992 see page 6, line 32 - page 7, line 32 see page 7, line 55 - page 8, line 3	1-13
P,Y	WO,A,9 119 782 (UNILEVER PLC) 26 December 1991 see page 7, line 32 - page 8, line 7; figure 1 see page 10, line 34 - page 11, line 12	1-13

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

EP 9202896
SA 68425

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 02/04/93

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-0421919	10-04-91	AU-A-	6205490	07-03-91
		JP-A-	3224489	03-10-91
EP-A-0357127	07-03-90	AU-A-	4002189	22-02-90
		JP-A-	2174673	06-07-90
EP-A-0463706	02-01-92	AU-A-	8318691	18-02-92
		AU-A-	8320591	18-02-92
		CA-A-	2067329	25-01-92
		EP-A-	0468596	29-01-92
		WO-A-	9201389	06-02-92
		WO-A-	9201793	06-02-92
WO-A-9119782	26-12-91	NL-A-	9001388	16-01-92
		AU-A-	8066891	07-01-92